Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, eds.

G-STAR: A NEW KRIGING-BASED TRUST REGION METHOD FOR GLOBAL OPTIMIZATION

Giulia Pedrielli
Szu Hui Ng

Department of Industrial & Systems Engineering
National University of Singapore
1 Engineering Drive, 4
SG 258127, SINGAPORE

ABSTRACT

Trust region methods are an efficient technique to identify good solutions when the sampling effort needs to be controlled due to the cost of running simulation. Meta-model based applications of trust region methods have already been proposed and their convergence has been characterized. Nevertheless, these approaches keep the strongly local characteristic of the original trust region method. This is not desirable in that information generated at local level are "lost" as the search progresses. A first consequence is that the search technique cannot guarantee global convergence. We propose a global version of the trust region method, the Global Stochastic Trust Augmented Region (G-STAR). The trust region is used to focus the simulation effort and balance between exploration and exploitation. Such an algorithm focuses the sampling effort in trust regions sequentially generated by adopting an extended Expected Improvement criterion. This paper presents the algorithm and the preliminary numerical results.

1 INTRODUCTION & PROBLEM STATEMENT

We consider a single objective minimization problem defined over a compact set $\mathbb{X}$. The deterministic d–dimensional objective function $f : \mathbf{x} \in \mathbb{X} \subset \Re^d \rightarrow f(\mathbf{x}) \in \Re$ can only be observed with noise from a simulator at any point $\mathbf{x}$. Our goal is to find a global minimum of $f : \mathbb{X} \rightarrow \Re$, where $\mathbb{X} \subseteq \Re^d$ solving:

$$\begin{aligned} P : \min & f(\mathbf{x}) \\ \text{s.to} \quad & \mathbf{x} \in \mathbb{X} \end{aligned}$$

In order to find a solution to P, we refer to meta-model based algorithms for simulation-optimization and we are specifically interested in two classes of approaches: (1) trust region methods producing locally converging procedures, and (2) meta-model based simulation optimization for global optimization. In fact, G-STAR bridges these two families of approaches aiming at providing to the class (1) a global convergence property. Such a link is established through the construction of ensembles of meta-models which are used to sequentially drive the search for the global optimum.

The Trust Region method (TR) was initially proposed as a deterministic nonlinear programming method with a similar framework to response surface methodology (RSM) (Nocedal and Wright 2006, Myers and Anderson-Cook 2009). At each iteration, the approach defines a trust region around the current centroid, it builds a quadratic approximation model, finds the optimal solution within the trust region (based on this quadratic model) and then moves to the optimal solution (if it is satisfactory by certain criterion). The size of trust region is automatically determined by the algorithm. In addition, it can be proved to converge to a stationary point of the objective function. Chang et al. (2013) proposed the stochastic version of this algorithm, the *Stochastic Trust Region Gradient-Free Method* (STRONG) for simulation optimization

with continuous decision variables. STRONG combines the traditional Response Surface Methodology framework with the trust region method for deterministic optimization to achieve convergence property and it provides an automated procedure to implement RSM. STRONG is proved by the authors to have the potential of solving high-dimensional problems efficiently. The results in the paper are particularly interesting. Nevertheless, similar to the deterministic case, information generated locally in terms of response surfaces, is only used to make local decisions, hindering the possibility to search for global solutions.

On the other hand, meta-model based global methods exploit the information coming from the simulation by iteratively improving the estimate of a *global* model of the response surface to optimize. Since this response surface is constructed upon a simulation model, it is referred to as *meta*–model. These procedures typically use the meta–model to compute an indicator (e.g., the Expected Improvement, Knowledge Gradient) which guides the search, by iteratively suggesting the next point(s) to sample with the objective to improve the meta-model and eventually identify the global optimum. The Efficient Global Optimization Algorithm (EGO) represents one of the most adopted procedures in this family for the case of deterministic $f(x)$. It uses the *Expected Improvement*, to choose the next point(s) and a ordinary kriging model to fit the response. Sequential Kriging Optimization (SKO) extends EGO to the stochastic case by adapting both the meta–model form as well as the criterion to guide the search. TSSO and, more recently, eTSSO, extend SKO by providing advanced sampling and budget allocation criteria as well as the possibility to handle non-homogeneous variance (Quan et al. 2013, Pedrielli and Ng 2015, Pedrielli and Ng 2016).

Meta-model-based approaches provide a way to model the global information about the response surface, thus leading to global optima. For this reason it is appealing to try integrate them in a TR framework. However, it is not straightforward to integrate the global meta-model into the trust region scheme. This is because, in the trust region approach, a local model is generated *within* the current visited region. As a result, the search progresses by generating local models and using a unique model as in the meta-model literature would contradict the TR approach. A way to integrate the global information with multiple local models is therefore necessary to exploit the TR framework, while maintaining a global information.

In this regard, meta-model ensembles represent an attractive alternative to be able too mix globally and locally generated models as it is our target (Goel et al. 2007, Zerpa et al. 2005). In this context, the meta-model ensemble is similar to model averaging, which serves as an approach to account for model uncertainty. Goel et al. (2007), Müller and Piché (2011) and Goel et al. (2007) suggest different approaches for determining the weights of models in the combination. In order to apply ensembles to guide the search in a TR based algorithm, we need to modify two main aspects. Firstly, in TR approaches, models are dynamically generated as the search progresses along with the related trust regions and they are not known in advance and, hence, to adapt to this, the ensemble model is developed and estimated dynamically. Secondly, the weights, computed at each iteration in the optimization procedure, need to be defined as a function of the location $\mathbf{x}$, and hence, an optimization problem needs to be solved at each iteration which assigns to each model a different weight based on the fitting performance.

G-STAR, as a global meta-model based method, iteratively updates a unique global model which helps in identifying centroids of trust regions to exploit. Subsequently, as a TR-method, iteratively builds meta-models within the generated trust regions. The generated models are subsequently integrated forming the *meta-model ensemble* which is used to predict the best point at each iteration. Specifically, the ensemble is constructed with weights associated to each model that reduce to 0 exponentially fast as we get farther from the model trust region.

2 THE G-STAR ALGORITHM

In order to provide a global feature to the trust region algorithm, G-STAR makes use of kriging as meta-model. This choice is due to the flexibility of kriging models in representing different shapes and, in the scope of optimization, the presence of several indicators for sampling, which we use to support the meta-model based search.

G-STAR is developed around three main routines: (1) initialization, (2) search, and (3) ensemble construction. An outline of the algorithm is provided in Figure 1.

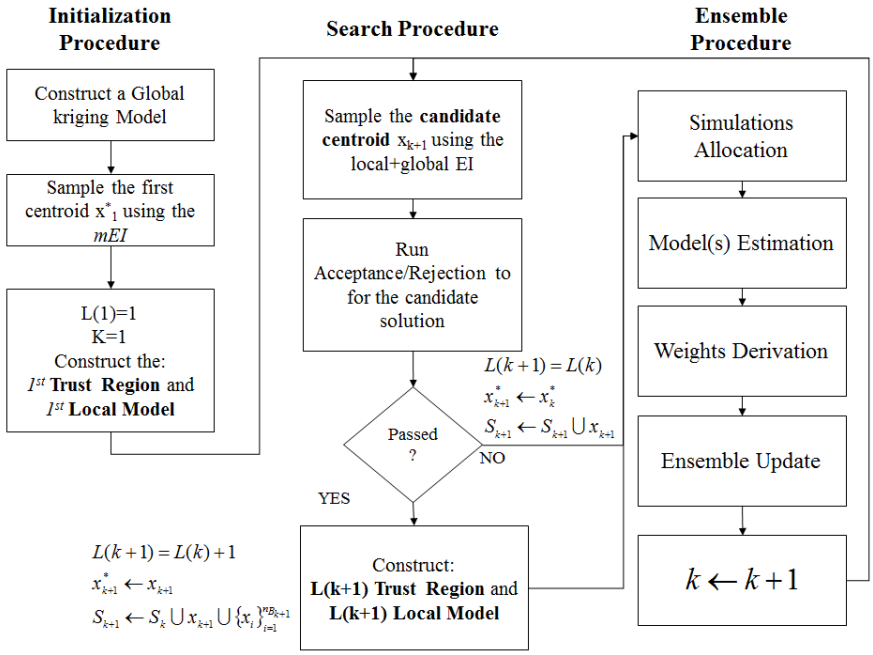


Figure 1: G-STAR Overview.

The main steps of G-STAR are outlined below:

1. **Initialization Procedure**:
   - Construct a first Global Model with n_0 sampled points from $\mathbb{X}$;
   - Use the Modified Expected Improvement (mEI), first proposed in (Quan et al. 2013), to derive the first centroid $\mathbf{x}_1^*$;
   - Construct a trust region of size Δ_0 and estimate the first local model. Set the number of iterations to $k = 1$, and update the number of trust regions at iteration k, $L(k) = 1$.
2. **Search Procedure**
   - Sample the centroid of the next candidate trust region, $\mathbf{x}_{k+1}$, according to the Trust Region expected improvement criterion proposed in this paper. This criterion samples either inside the last created trust region or in the feasible region $\mathbb{X}$ excluding the already sampled points in the set $\mathbb{S}$ as well as the generated trust region.
   - Evaluate the function $\bar{f}(\mathbf{x}_k)$ by running m_{min} simulation replications at that new candidate centroid $\mathbf{x}_{k+1}$;
   - Conduct the modified ratio-comparison and sufficient-reduction tests in order to examine the current $\mathbf{x}_{k+1}$.
     - In case the candidate is accepted, update the centroid of the trust region $\mathbf{x}_{k+1}^*$ and the size of the trust region $\Delta_{L(k+1)}$. A new local model is created by uniformly sampling $n_{\mathscr{B}_{k+1}}$ points within the trust region. The set of sampled points at iteration $k+1$, S_{k+1}, is updated by adding the centroid $\mathbf{x}_{k+1}$ and the uniformly sampled points $\{\mathbf{x}_i\}_{i=1}^{n_{\mathscr{B}_{k+1}}}$. The number of trust regions is updated $L(k+1) \leftarrow L(k)+1$;
     - If the tests are not passed, then the trust region is reduced in size and the local model is updated considering the new sampled point.

3. **Ensemble Procedure**
   - According to the Ensemble procedure, the simulation budget remaining from the search step is allocated to the already sampled points and the local model(s) as well as the global response are re-estimated accordingly. This means that *all* local models and the global model can be potentially re-estimated any time a point within the trust region or a centroid receives additional simulation budget. Specifically, $W_{i,k}$ represents the number of simulation replications allocated to point i up to iteration k. We will use $m_k = W_k - W_{k-1}$ and $m_{ik} = W_{ik} - W_{i,k-1}$, $i \in \mathbb{S}$, to refer to the number of simulation replications allocated at iteration k and the number of simulation replications allocated to point i at iteration k, respectively.
     As a result, we can produce the local forecasts for each trust region $\hat{f}_i^{\ell}(\mathbf{x}_0)$, $i = 1, \ldots, L(k+1)$ as well as the global prediction $\hat{f}^g(\mathbf{x}_0)$;
   - We use the produced predictions to compute the ensemble of the meta-models by deriving the weights $w_g(\mathbf{x}; \theta)$ for all the models.

Given a total simulation budget T, the procedure stops when the used budget is $W_k = T$. In order to run G-STAR, an initial size of the trust region Δ_0 needs to be provided as input parameter and used every time a new trust region is generated. At the end, the sampled point with the lowest mean is selected.
In the following part of the paper, we will go into the details of the algorithm components starting form the *Ensemble Procedure*, section 2.1, whereas the *Search Procedure* will be further detailed in section 2.2.

2.1 Estimation of the Ensemble of Surrogates

In this section, we present the modeling technique applied in this paper. In particular, we aim at generating a meta-model ensemble by sequentially assigning weights to the different meta-models, developed in the different trust regions, proportionally to the Mean Squared Error estimated throughout the sampled space.

In the context of this paper, the set of meta-models is not known beforehand and, instead, we dynamically update a global surrogate model by using the function evaluations at the sampled centroids, while we estimate a local surrogate model in each of the trust regions being generated around the centroids. As it will be further detailed in section 2.2, the centroids are sampled according to the *expected improvement* criterion. As a result, similarly to the literature on meta-model based simulation-optimization, we interpret these points as representative of the global response characteristics. In fact, they are sampled in the scope of exploring and exploiting the search space.

The idea behind this approach is that, while the global model gives us a chance to explore the entire solution space, the local models focus on locally improving the global fit by considering only the points within the trust regions and neglecting other sampled regions. As a result of this, we expect that the local kriging model will be particularly accurate within the trust region but not outside. This consideration is relevant in the scope of computing the weights to assign to each estimated meta-model, as it will be in the next paragraphs.
As a result of the weighted model update, the response prediction can be formulated as it follows ($L(k)$ represents the number of trust regions created up to iteration k):

$$\hat{f}^e(x) = \sum_{i=1}^{L(k)} w_i(\mathbf{x}; \theta)\, \hat{f}_i^{\ell}(\mathbf{x}) + \left(1 - \sum_{i=1}^{L(k)} w_i(\mathbf{x}; \theta)\right) \hat{f}^g(\mathbf{x}), \quad \mathbf{x} \in \mathbb{X}. \tag{2.1}$$

Where $\hat{f}^e(x)$ represents the ensemble prediction at the generic point x in the feasible space, and it is derived based on $\hat{f}_i^{\ell}$, referring to the local models estimated using the points sampled in the i–th trust regions and $\hat{f}^g(x)$ estimated using the centroids generated throughout the algorithm iterates.
Despite the local models and the global model are estimated according to different sampled points, the fit is always global and this strategy is used in order to simplify the construction of a smooth ensemble of

surrogates. Indeed, even if the prediction is global, local models will be given a very low weight outside their TR, while still guaranteeing smoothness of the overall prediction. Specifically, all these predictions, following the meta-models ensemble literature, are averaged by means of a series of weights $w_i(\mathbf{x};\theta)$ which are a function of the points location $\mathbf{x}$ and parametrized over θ, which controls the rate at which the local model weight vanishes. It has to be noted that, as highlighted in Figure 1, at the end of the search procedure we may not generate a new trust region. For this reason, generally, $L(k) \neq k$ and $L(k+1) \geq L(k)$.

One peculiar aspect for the prediction proposed in equation (2.1) is that we treat the global model differently from the local surrogates. The main reason is that the global model is estimated independently from the local surrogates in that it only uses the function values at the centroids (which are not used to estimate the trust region models). The choice of using centroids, sampled according to the *global* expected improvement (equation (2.5)), is in line with the traditional literature on meta-model based global optimization, as already mentioned. In fact, the main role of the global model in G-STAR is to guarantee global exploration of the solution space which is realized through the generation of centroids, as it will be clearer in section 2.2.1. In fact, the global response enables us to consider the entire feasible space even when no local model is defined. On the other hand, the availability of meta-models enables the computation of sampling criteria at global level, thus avoiding local minima as typical of the original trust region algorithm.

In the next part, we briefly describe the base meta-model used by G-STAR.

Models Estimation Differently from the traditional literature on meta-models ensembles, in this paper each model adopted is in the family of Modified Nugget Effect Kriging (MNEK) model (Yin et al. 2011). The choice of this meta-model is motivated by the stochastic nature of simulation and non-homogeneous variance. As previously mentioned, the choice of kriging models was performed for the flexibility in terms of shapes that can be produced if compared to polynomial forms. At global level, the choice of a kriging model is necessary in the scope of capturing a global trend. On the other hand, at local level, simpler models might also be considered. In particular, as in the tradtional TR literature, first and second order models might be locally estimated. The proposed approach does not lose in generality assuming different local models. However, Gaussian models represent an advantage in view of developing a closed form estimation for the MSE of the meta-model ensemble.

We assume that the function values $f(\mathbf{x})$ are realizations of a random process and a statistical model represents the response surface, namely:

$$f(\mathbf{x}) = Z(\mathbf{x}) + \xi(\mathbf{x}), \quad \mathbf{x} \in \mathbb{X} \tag{2.2}$$

where Z describes the mean and ξ describes the random noise process. As in the deterministic case, we further model $Z(\mathbf{x})$ as a Gaussian process with covariance function $\tau^2 V$, where τ^2 is the process variance and V the matrix of process correlation; formally, $Z(\mathbf{x}) \sim GP\left(\mu(\mathbf{x}), \tau^2 V\right)$.

A commonly adopted correlation function $V = \left(K_\theta(x_i - x_j)\right)_{i=1}^{k}$ is the d–dimensional separable version of the power exponential family of functional forms which is characterized by smooth response. The noise $\xi(x)$ is assumed to be distributed with zero mean and covariance function $\sigma_\xi^2 V_\xi$, where V_ξ denotes the matrix of *sampling* correlations. Error variances are generally not constant and they may depend on $\mathbf{x}$ (i.e., the heteroscedastic case is considered). With independent sampling (i.e., no CRN), V_ξ is diagonal, and equation (2.2) reduces to the independent sampling noise model Yin et al. (2011). The general form of equation (2.2) is similar to the form proposed in Ankenman et al. (2010).

As shown in Yin et al. (2011), the predictor for (2.2) at the point x, given k points have been already sampled, is:

$$\hat{f}(m_k, x, \theta) = \sum_{i=1}^{k} \left(v^T \left(V_\xi + V\right)^{-1} e_i + 1^T \left(V_\xi + V\right)^{-1} \frac{\left[1 - 1^T \left(V_\xi + V\right)^{-1} v\right]^T}{1^T \left(V_\xi + V\right)^{-1} 1} e_i \right) \bar{f}_i \tag{2.3}$$

where, m_k represents the total number of simulation replications performed up to iteration k. $\bar{\mathbf{f}}$ is the k–dimensional vector of the averaged function values at the already sampled points. v is the correlation vector, specifically, $v(m_k,x,\nu)^T = \left(e^{-\nu\cdot d^2_{x,x_1}} \cdots e^{-\nu\cdot d^2_{x,x_k}}\right)$, where, d_{x,x_i} represents the euclidean distance between point x at which the prediction is performed and the already sampled locations x_i, $i = 1,\ldots,k$. The vector e_i has size k (being k the number of sampled points) and its elements are all equal to 0 except the i–th element which is equal to 1.

Weights Derivation Since G-STAR dynamically generates local models, each iteration will potentially have a different number of meta-models. Since the local models are meant to be fit for the specific trust region, intuitively, they should receive a larger weight within the related trust region and such a weight should vanish to 0 as the distance from the region increases (in terms of distance from the centroid). In light of these observations, we propose to weight each meta-model according to the following form:

$$w_i(\mathbf{x};\theta) = \exp\left(-\theta_i \frac{\sum_{l=1}^{d}(x_l - r_{i,l})^2}{2|\mathscr{B}_i|}\right),\ i = 1,\ldots,L(k) \tag{2.4}$$

Here, $L(k)$ refers to the number of trust regions at iteration k, the parameters $\{\theta_i\}_{i=1}^{i=L(k)}$ will be estimated by minimizing the MSE of the ensemble and they represent an additional degree of freedom to optimize the rate of decay of the weight in the space. The vector $\mathbf{r}_i$ represents the d-dimensional centroid of the trust region and $\mathbf{x}$ is the generic point being considered. Finally, $\mathscr{B}_i$ represents the trust region and $|\cdot|$ refers to the size of the trust region. The inspiration of the weight comes from the Gaussian density, where the deviation is interpreted as the size of the trust region.

2.2 Search Procedure

2.2.1 The Trust Region-Expected Improvement ($EI^{\mathscr{B}}_{\tilde{\pi}_k}$)

At each iteration k a global response model and a new local response model are both constructed. We use them as support to identify the next trust region centroid as it is common in a trust–region approach. Specifically, at iteration k, we consider the current centroid x_k, the corresponding trust region $\mathscr{B}_{L(k)}\left(x_k,\Delta_{L(k)}\right)$ of size Δ_k, the global response model estimated with the sampled centroids and the last estimated local model over the current trust region.

In order to find the next candidate centroid, we propose the trust region-modified Expected Improvement indicator ($EI^{\mathscr{B}}_{\tilde{\pi}_k}$), the indicator is formulated as follows:

$$EI^{\mathscr{B}}_{\tilde{\pi}_k} = 1_{g\left(\Delta_{L(k)}\right)\leq 0} EI_{\tilde{\pi}_k}(\mathbb{X}) + 1_{g\left(\Delta_{L(k)}\right)>0} EI_{\tilde{\pi}_k}(\mathscr{B}) \tag{2.5}$$

In the above equation, $EI_{\tilde{\pi}_k}$ is defined in a different way based on the fact that x^*_{k+1} will be sampled within or outside the trust region. In fact, due to the construction of the algorithm, the local and global criterion are exclusive and the activation of one of the two is guided by the indicator function $1_{g\left(\Delta_{L(k)}\right)\geq 0}$. We refer to the component $EI_{\tilde{\pi}_k}(\mathbb{X})$ in equation 2.5 as *Global EI* , whereas the *Local EI* is $EI_{\tilde{\pi}_k}(\mathscr{B})$.

We will sample according to the global (or local) EI when the following function is less (larger) than or equal to 0:

$$g\left(\Delta_{L(k)}\right) = \left[\left(\frac{\Delta_{L(k)}}{\Delta_0} + \varepsilon\right) - 1\right] \tag{2.6}$$

Where, ε is a random variable (a uniform $\mathscr{U}(0,1)$). We can notice that this criterion guarantees that when the size of the trust region $\Delta_{L(k)} \to 0$, then the function $g \leq 0$ no matter the realization of ε. Therefore a

point for the next centroid will be created outside the current trust region. The search is then brought back by construction to the global search typical of meta-model based simulation optimization, thus guaranteeing the global convergence.
When $\Delta_{L(k)} > 0$, instead, the probability of sampling inside the trust region is proportional to the size of the trust region with respect to the initial size Δ_0 (which is an input parameter of the algorithm). The idea is that, at the beginning we have large probability to sample inside the trust region and this probability decreases with the iterations (as the trust region reduces in size). In case the sampling happens outside the trust region, i.e., $g\left(\Delta_{L(k)}\right) \leq 0$, then the criterion is equivalent to the one in (Pedrielli and Ng 2015) and the EI is computed considering the entire feasible space excluding the current trust region, $\mathscr{B}_{L(k)}$, and the set of already sampled points:

$$x_{k+1} \in \arg \max_{x \in \mathbb{X} \notin \left(\mathbb{S} \cup \mathscr{B}_{L(k)}\right)} EI_{\tilde{\pi}_k} := \mathbb{E}_{\tilde{\pi}_k}\left[\max\left\{\bar{f}\left(\mathbf{x}_k^*\right) - \hat{f}^g\left(\mathbf{x}\right), 0\right\}\right] \tag{2.7}$$

In equation (2.7), $\mathbb{X}$ represents the entire feasible space and $\mathbb{S}$ the set of already sampled points. x_k^* refers to the point with the best function value $\bar{f}^g\left(x_k^*\right)$ up to iteration k, while $\hat{f}\left(x\right)$ is the predicted global function value at the generic point $\mathbf{x}$.

When the point is sampled inside the trust region, the same expected improvement is computed, but *within the trust region*, i.e., $\mathbb{X} \leftarrow \mathscr{B}_k$. In this sense, we look for the best point within the trust region since we believe it is a promising region, and we use the expected improvement locally. We highlight that several approaches may be used in order to search locally. Nevertheless, the expected improvement gives us the possibility to leverage on the produced local model when sampling, making use of the Gaussian process definition to derive the indicator.

In general, it can be argued that the Expected Improvement is a criterion which leads to both exploration and exploitation, which means that the EI may choose to sample very close to a previous trust region or inside an already existing trust region even when the condition $g\left(\Delta_k\right) \leq 0$ is verified and we would like to get out from the trust region. Nevertheless, it is relevant to highlight that the *global EI* looks for new centroids and the presence of previously generated trust regions where the global EI may sample does not contradict the aim of the global search for which G-STAR is proposed. Simply, overlapping trust regions will be generated. Also, the local and global sampling criteria will never choose a point from the same set. This guarantees G-STAR to reach density in the sampling space as the number of iterations $k \to \infty$. Indeed, by construction, the algorithm will never choose the same centroid twice and, on the other hand, the condition $g\left(\Delta_{L(k)}\right) \leq 0$ will guarantee exhaustive search in the sampling space.

By construction, the following property holds:

Property 2.1. *For any iteration k, we have:* $\hat{f}_k^e\left(\mathbf{x}_k\right) - \hat{f}_k^e\left(\mathbf{x}_{k+1}\right) \geq EI_{\tilde{\pi}_k}^{\mathscr{B}}$*, where* $\hat{f}_k^e$ *is the prediction computed from the meta-model ensemble and* $\mathbf{x}_{k+1}$ *represents the candidate centroid, i.e., the point solving 2.5.*

The property can be proved just considering the definition of expected improvement in equation (2.7). This fact will be particularly useful in the derivation of the sufficient reduction test to establish whether to move the centroid of the current trust region at the generic iteration k.
The proposed sampling criterion is one of the differences between G-STAR and STRONG, which looks for a Cauchy point in the local region and does not provide a global search criterion. Due to this characteristic of G-STAR, the algorithm can reach global convergence. Indeed the global criterion will always be active when the size of the trust region goes to 0. This enables us to use the convergence results in the previous work of the authors (Pedrielli and Ng 2016).

2.2.2 Acceptance/rejection of the new centroid

As a TR algorithm, G-STAR automatically generates a new centroid/ updates the trust region based on the results from the Ratio-Comparison (RC) and Sufficient Reduction SR tests. Despite in the original paper (Chang et al. 2013), the authors do not use as response surface a kriging model, it can be argued that the two tests play the same role in G-STAR as in STRONG: the RC test verifies the consistency between the

prediction and the simulated values and this is consistent with the idea of model variance in the kriging literature. The sufficient reduction test, compares the actual difference between two consecutively sampled points and the expected difference, which is the expected improvement in the context of meta-model based search.

The RC test fails when the simulated difference is much lower than the predicted difference between the function value at two consecutive centroids. The RC test is (Chang et al. 2013):$\rho_k = \frac{\bar{f}_k(\mathbf{x}_k)-\bar{f}_k(\mathbf{x}_{k+1})}{\hat{f}^e_k(\mathbf{x}_k)-\hat{f}^e_k(\mathbf{x}_{k+1})}$. Given two input parameters $0 < \eta_0 < \eta_1 < 1$, if $\rho_k < \eta_0$, then the RC test is *failed*. The prediction $\hat{f}_k$ here refers to the meta-model ensemble.
The only difference with respect to (Chang et al. 2013) here is that we are considering $\mathbf{x}_{k+1}$, i.e., the current centroid as a maximizer of the expected improvement instead of a Cauchy point. This is in line with our overall aim of making the trust region a global search algorithm. The Cauchy point is generated with gradient information, therefore it is subject to local convergence issues. Instead, the expected improvement is proven to converge to the global optimum. This does not modify the validity of the test in conceptual terms: since we are still using a meta-model to emulate the response surface, we are interested in verifying the quality of such a model.

On the other hand, even when the models are consistent, we might not update the centroid if the improvement is not significant. This is verified by the SR test. The sufficient reduction test is identical to the one proposed in (Chang et al. 2013): $t^* = \frac{\bar{f}_k(\mathbf{x}_k)-\bar{f}_k(\mathbf{x}_{k+1})-\eta^0\zeta_k}{S_k}$.
where S_k is the pooled sample variance of $\mathbf{x}_{k+1}$ and x_k and, by construction, $\zeta_k := EI^{\mathscr{B}}_{\tilde{\pi}_k}$, i.e., the value of the expected improvement. The variance adopted for the test is a pooled estimator: $S^2_k = \frac{S^2_k(\mathbf{x}_k,n_k)}{n_k} + \frac{S^2_k(\mathbf{x}_{k+1},n_0)}{n_0}$. The degrees of freedom of the resulting t–statistics for the adopted approach are:$df = S^4_k \cdot \left[\frac{\left(S^2_k(\mathbf{x}_k,n_k)/n_k\right)^2}{n_k} + \frac{\left(S^2_k(\mathbf{x}_{k+1},n_0)/n_0^2\right)^2}{n_0}\right]^{-1}$ It is important to notice that the test is not based on the predicted variance but the sampled variance at the point to be explored, where n_k represents the budget dedicated to the sampling of the point. The sufficient reduction test is passed in case $t^* > t_{1-\alpha_k,df}$ (Myers and Anderson-Cook 2009).

It has to be noted that, due to Property 2.1, we are able to translate the criterion originally presented in (Chang et al. 2013), to the case of sufficient reduction in equation (2.2.2) in G-STAR. This is due to the fact that we are able to bound the maximum expected improvement at each algorithm iteration and this is made possible by the use of meta-models.

2.3 Ensemble Update

After the centroid-tests, G-STAR uses the enlargement coefficient $\gamma_2 > 1$ and the shrinkage coefficient $0 < \gamma_1 < 1$ for modifying the size of the trust region. G-STAR will perform the following steps:

- If the candidate point passes the SR/RC Test, then the centroid is moved, i.e., $\mathbf{x}^*_{k+1} \leftarrow \mathbf{x}_{k+1}$ and the TR around the point is enlarged using $\Delta_{L(k+1)} = \gamma_2 \cdot \Delta_{L(k)}$. Once a trust region is determined $n_{\mathscr{B}_k}$ points are sampled within the trust region according to a Latin Hypercube sampling design in order to have a first estimate of the local model. Such a sampling choice is originates by the procedure we use to initially estimate a model in the surrogate literature, i.e., through the choice of n_0 initial sampling points Kleijnen (2008).
- If the point fails either RC or SR tests we have two scenarios:
  - If $\mathbf{x}_{k+1} \in \mathscr{B}$: The trust region is reduced according to $\Delta_{L(k+1)} = \gamma_1 \cdot \Delta_{L(k)}$ and the centroid is not updated.
  - if $\mathbf{x}_{k+1} \notin \mathscr{B}$: the solution is accepted, $\mathbf{x}^*_{k+1} \leftarrow x_{k+1}$, and the trust region is enlarged (favor exploration), i.e., $\Delta_{L(k+1)} = \gamma_2 \cdot \Delta_{L(k)}$.

It is important to highlight that the shirkage of trust region in G-STAR does not lead to the construction of a *new* meta-model. A result of the shrinkage is the increase in the sampling density (in terms of ratio between sampled points and size of sampling space) and this leads intuitively to an improved fit of the kriging model. It may be argued that simply increasing sampling in the trust region without reducing the size will lead to the same effect of improving the model fit, due to the properties of kriging. However, reducing the size of the trust region can have the positive effect of increasing the speed of exploration of the space by reducing the probability of sampling within the TR (equation 2.5). Nevertheless, it is has to be noted that, wile TR reduction in the original algorithm is *required* to fit the local model, in our case it just helps in increasing exploration.

Sequential Update of the Meta-model Ensemble The sequential update of the response surface happens in two steps: (1) update of the global and local model(s), (2) update the set of sampled points and subsequent update of the weights. Specifically, at each iteration of the algorithm we estimate the parameters of:(A) The global response surface considering only the centroid(s) generated by the global component of the Expected improvement in equation (2.7). This implies that, in case the sampling is local, then no estimation of the global parameters is performed; (B) Generate $n_{\mathscr{B}_{L(k)}}$ Latin Hypercube samples in the trust region $\mathscr{B}_{L(k)}$ and update the estimate of *all* the local models having trust regions intersecting with the set of new sampled points. This results in an improvement of multiple local models every time a sampling region is formed according to the latin hypercube sampling as described in section 2.3. For the update of the overlapping trust regions, we always consider the initial size Δ_0.; (C) Evaluation: once the new points in the trust region have been sampled, we may have remaining budget for the iteration to improve the model estimation. Specifically, the total budget to allocate to the evaluation stage is established according to the criterion presented in (Pedrielli and Ng 2015, Pedrielli and Ng 2016). In particular, allocating budget is important for two reasons: (1) computational effort to dedicate to a specific iteration of the algorithm, (2) computational effort to dedicate to each sampled solution to improve the response estimate. (Pedrielli and Ng 2015) solves (1), while Optimal Computing Budget Allocation (OCBA) scheme is used for (2). After simulation, all involved trust regions and the global model may be re-estimated; (D) Recompute the optimal weights for all the regions solving the MSE problem in (2.8).

We propose to minimize the empirical MSE to define the weights to associate to the different models. In particular, at each iteration, *all* the weights parametrized by θ in equation (2.8) at each local region and, consequently, the weight of the global model are updated by solving the following optimization problem:

$$\min_{\theta} MSE_k(\theta) = \min_{\theta} \frac{1}{|\mathbb{S}|} \sum_{\mathbf{x}\in\mathbb{S}} \left(\hat{f}_k^e(\mathbf{x}) - \bar{f}(\mathbf{x})\right)^2 =$$

$$\frac{1}{|\mathbb{S}|} \sum_{\mathbf{x}\in\mathbb{S}} \left(\left[\sum_{i=1}^{L(k)} w_i(\mathbf{x};\theta)\, \hat{f}_i^{\ell}(\mathbf{x}) + \left(1 - \sum_{i=1}^{L(k)} w_i(\mathbf{x};\theta)\right) \hat{f}^{g}(\mathbf{x}) \right] - \bar{f}(\mathbf{x}) \right)^2 \tag{2.8}$$

When the global prediction is pretty good, we will expect θ in the weighting function to grow in order to taper down the effect of the local model very quickly around the centroid in order to let the global model play the major role in the prediction. It is apparent that the MSE problem just formulated will increase its computational demand when the iterations k and, consequently, the number of trust regions $L(k)$ remarkably grow.

It is also noteworthy, that multiple local models are updated along the way, and this may be computationally expensive as the number of local models increases with the number of iterations. Updating schemes which consider the likelihood of a trust region to contain the global optimimum in

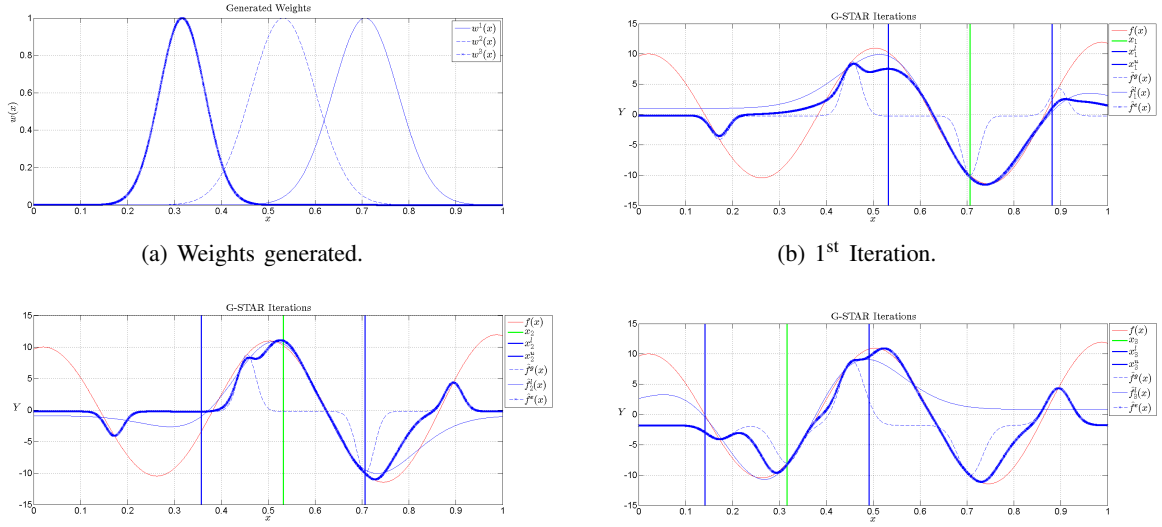
3 PRELIMINARY NUMERICAL RESULTS

In this section, we show the behavior of G-STAR over a simple 1–dimensional case. Since G-STAR introduces a global characterization to trust region based approaches, we are interested in investigating the workings of the algorithm and its empirical properties in terms of convergence to the global optimum for a

multi-modal response surface as the one in the example. In order to test the performance of the algorithm, we compare G-STAR with another global algorithm developed by the authors, i.e., eTSSO (Pedrielli and Ng 2016). In fact, since TR algorithms are not able to guarantee global convergence, a more fair comparison is between G-STAR and globally converging procedures. Specifically, we used the following function as test example in this paper:

$$f(x) = (2x + 9.96)\cos(13x - 0.26) \tag{3.1}$$

This function has a global minimum in $x^* = 0.746$ with function value $y^* := y(\mathbf{x}^*) = -11.45$ and a local minimum in $x = 0.2628$, with $\mathbb{X} = [0,1]$. As noise, we applied to the function an additive Gaussian Process $\xi(x)$ with mean $\mathbf{0}$ and diagonal variance covariance matrix with diagonal elements (δ represents the magnitude of the noise): $\sigma^2_\xi(x) = \delta \cdot x$. Figure 2 shows three sequential iterations of G-STAR. In particular, a set of $n_0 = 3$ points are sampled in $\mathbf{x}_0 = [0.4551, 0.61723, 0.9951]$ and a first global model estimation is performed. With this first global model, the first centroid $x = 0.7072$ for the trust region with size $\Delta_0 = 0.3499$ (where Δ_0 is an input parameter). Figure 2b shows the fitting of the first local model in the trust region defined by the centroid x^*_k and the lower and upper bounds $\left(x^l_1, x^u_1\right)$. As expected, the local model fits particularly well the true function within the trust region, but poorly outside of its estimation interval. As the algorithm progresses (Figures 2c-2d), we notice that the meta-model ensemble is incrementally improved due to the generation of new local models. The proposed methodology for ensembles produces a surrogate which basically conforms with the local model in the trust region, while following the global model outside it. We can notice that, at the 2^{nd} iteration no update of the global model



(a) Weights generated.

(b) 1st Iteration.

(c) 2nd Iteration.

(d) 3rd Iteration.

Figure 2: Sequential Ensemble Construction.

is performed since the new centroid is generated by means of the local EI (2.5). Moreover, since the trust region changes at each iteration, we do not observe any shrinkage.

G-STAR looks particularly promising when low budget values are considered. Indeed, in such a case, the algorithm can effectively take advantage of a local good model to improve the global one, whereas the original optimizer has to deal with a low quality global model. It is important to notice that, as shown in the figures, when the number of iteration progresses, the global model will tend to have more weight that any local model. This is asymptotically correct due to the fact that, as the iterations go to infinity we will sample, by construction, an infinite number of centroids. Therefore, the global model will be dense in

Common Parameters		G-STAR Parameters	
n_0	3	γ_1	0.6
m_{min}	10	γ_2	1.05
T	200	η_0	0.2
δ	0.1	η_1	0.8
		α	0.05
		$n_{\mathscr{B}_k}$	4

(a) Algorithms Parametrization

	eTSSO	GSTAR
Average $\lvert x-x^*\rvert$	0.45	0.0194
Std Err $\lvert x-x^*\rvert$	0.0122	0.021
Average $\lvert y-y^*\rvert$	0.919	0.433
Std Err $\lvert y-y^*\rvert$	0.019	0.046
PCS	7%	66%

(b) Comparison between eTSSO and G-STAR

the solution space. Considering the local models being generated, as the iterations $k \to \infty$, the optimized parameter θ will tend to become larger to reduce the influence of the local model.

In this preliminary analysis, we had as target to understand the behavior of the algorithm, particularly in the way it sequentially estimates the response surface, and compare it with the previous algorithm designed and extended by the authors, i.e., the Two Stage Sequential Optimization algorithm (TSSO).

In order to run TSSO and G-STAR we adopted the parametrization in Table 1a. Therefore, $n_{\mathscr{B}_k} = 4$ points are used to estimate the local models.

In the first set of results, we show the workings of the algorithm in terms of response surface estimation. In particular, we show how progressively the ensemble of surrogates fits the true function.

In order to compare the performance of G-STAR and eTSSO (Pedrielli and Ng 2016), we fixed the same initial points and we adopted the same random number stream throughout the algorithm execution. We compare the two algorithms based on three KPI's: The *location error*, i.e., $|x-x^*|$ computed as the euclidean distance between the solution produced by the algorithms and the true global optimum. We also report the error $|y-y^*|$ which measures the precision in the estimation of the value of the function at the global optimum. Moreover, we report the results in terms of *Probability of Correct Selection* (PCS) which is computed as the percentage over the 100 macro-replications, that the algorithm chooses the global optimum instead of the local solution in $x = 0.2628$. Table 1b shows the performance of the two algorithms.

It is apparent how, under a very limited budget, as it is in the case of the experiment performed, G-STAR shows a statistically significant improvement in the function estimation and, as a result, in the identification of the global minimum.

4 CONCLUSIONS

This paper proposed the Global Stochastic Trust Augmented Region Algorithm for the first time. The main idea of G-STAR is to extend trust-region methods to provide them a global optimization perspective, while contributing in the area of ensemble of surrogates by providing a new way to dynamically assign location-dependent weights which represents a novelty in the surrogate literature where models are typically weighted by constant factors all through the response surface. The development of the algorithm highlights the issues in sampling, where local and global conditions need to be verified when generating a candidate point. In order to do so, we create an index which guarantees us to escape local solutions. This extension of the Expected Improvement criterion is particularly suitable when meta-models are available as it makes explicit use of them. The preliminary numerical analysis shows that, when the budget is relatively small, G-STAR provides a structural advantage with respect to the algorithms previously developed by the authors, thus making the further development of the technique an interesting topic to investigate, especially in higher dimensions. Different local model types can be tested of simpler natural such as second or first order models. When Gaussian models are used, then G-STAR might take particular advantage of the closed form derivation of the MSE of the meta-model ensemble. The finite time performance analysis should be performed to provide theoretical strengths of the ensembles approach. Also, several parameters need to be tuned, adaptive versions of G-STAR should be studied to avoid their manual setting.

REFERENCES

Ankenman, B., B. L. Nelson, and J. Staum. 2010. "Stochastic kriging for simulation metamodeling". *Operations research* 58 (2): 371–382.

Chang, K.-H., L. J. Hong, and H. Wan. 2013. "Stochastic trust-region response-surface method (strong)-a new response-surface framework for simulation optimization". *INFORMS Journal on Computing* 25 (2): 230–243.

Goel, T., R. T. Haftka, W. Shyy, and N. V. Queipo. 2007. "Ensemble of surrogates". *Structural and Multidisciplinary Optimization* 33 (3): 199–216.

Kleijnen, J. P. C. 2008. *Design and analysis of simulation experiments*, Volume 111 of *International Series in Operations Research & Management science*. Springer.

Müller, J., and R. Piché. 2011. "Mixture surrogate models based on Dempster-Shafer theory for global optimization problems". *Journal of Global Optimization* 51 (1): 79–104.

Myers, R. H., and C. M. Anderson-Cook. 2009. *Response surface methodology: process and product optimization using designed experiments*, Volume 705. John Wiley & Sons.

Nocedal, J., and S. Wright. 2006. *Numerical optimization*. Springer Science & Business Media.

Pedrielli, G., and S. H. Ng. 2015. "Kriging-based Simulation Optimization: A Stochastic Recursion Perspective". In *Proceedings of the 2015 Winter Simulation Conference, L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.*, 3834–3845. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Pedrielli, G., and S. H. Ng. 2016. "Two Stage Sequential Optimization Approach: Generalization and Asymptotic Properties". *Submitted to IIE Transactions. Under Review*.

Quan, N., J. Yin, S. H. Ng, and L. H. Lee. 2013. "Simulation optimization via kriging: a sequential search using expected improvement with computing budget constraints". *Iie Transactions* 45 (7): 763–780.

Yin, J., S. H. Ng, and K. M. Ng. 2011. "Kriging metamodel with modified nugget-effect: The heteroscedastic variance case". *Computers & Industrial Engineering* 61 (3): 760–777.

Zerpa, L. E., N. V. Queipo, S. Pintos, and J.-L. Salager. 2005. "An optimization methodology of alkaline–surfactant–polymer flooding processes using field scale numerical simulation and multiple surrogates". *Journal of Petroleum Science and Engineering* 47 (3): 197–208.

AUTHOR BIOGRAPHIES

GIULIA PEDRIELLI is Research Fellow for the Department of Industrial & Systems Engineering at the National University of Singapore. Her research focuses on stochastic simulation-optimization in both single and multiple–objectives framework. She is developing her research in meta-model based simulation optimization and learning for simulation and simulation optimization. Her email address is giulia.pedrielli.85@gmail.com.

SZU HUI NG is an Associate Professor in the Department of Industrial and Systems Engineering at the National University of Singapore. She holds B.S., M.S. and Ph.D. degrees in Industrial and Operations Engineering from the University of Michigan. Her research interests include computer simulation modeling and analysis, design of experiments and quality and reliability engineering. She is a member of IEEE and INFORMS, and a senior member of IIE. Her email address is isensh@nus.edu.sg and her website is http://www.ise.nus.edu.sg/staff/ngsh/index.html.