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.

KRIGING–BASED SIMULATION–OPTIMIZATION: A STOCHASTIC RECURSION PERSPECTIVE

Giulia Pedrielli

Centre for Maritime Studies
National University of Singapore
15 Prince George's Park
Singapore, SG 118414, Singapore

Szu Hui Ng

Department of Industrial and Systems Engineering
National University of Singapore
1 Engineering Drive 2
Singapore, SG 117576, Singapore

ABSTRACT

Motivated by our recent extension of the Two–Stage Sequential Algorithm (eTSSO), we propose an adaptation of the framework in Pasupathy et al. (2015) for the study of convergence of kriging–based procedures. Specifically, we extend the proof scheme in Pasupathy et al. (2015) to the class of kriging–based simulation–optimization algorithms. In particular, the asymptotic convergence and the convergence rate of eTSSO are investigated by interpreting the kriging–based search as a stochastic recursion. We show the parallelism between the two paradigms and exploit the deterministic counterpart of eTSSO, the more famous Efficient Global Optimization (EGO) procedure, in order to derive eTSSO structural properties. This work represents a first step towards a general proof framework for the asymptotic convergence and convergence rate analysis of meta–model based simulation–optimization.

1 INTRODUCTION

In the last decade, simulation has been adopted as a means to iteratively evaluate the performance of the solutions generated by search algorithms (Fu et al. 2008). In the literature, this coupling is referred to as simulation–optimization. This approach has been proven particularly effective when complex functions are of concern, and the search algorithm has to rely upon the performance estimation produced by the simulation (typically treated as a black box oracle) at selected points (candidate solutions) since a closed form of the function is not available (Fu et al. 2008).

In this work, we specifically refer to the family of simulation–optimization problems trying to find the point x^* satisfying $x^* \in \arg\min_{x\in\mathbb{X}} f(x)$, where $\mathbb{X}$ represents the design space (i.e., the feasible region of the problem), and $f(x)$ is the function whose measurements (with or without noise) can be obtained only through simulation. More specifically, the scope of this manuscript is the one of meta–model based simulation–optimization and, in particular, we will focus on kriging–based simulation–optimization.

The basic idea behind this family of algorithms is to exploit the information coming from the simulation by iteratively improving the estimate of a 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, iteratively suggesting the next point(s) to sample with the objective to identify the global optimum.

The Efficient Global Optimization Algorithm (EGO) represents one of the most famous of this family of procedures when the function $f(x)$ is deterministic. It uses the *Expected Improvement*, to choose the next point(s) and a ordinary kriging model to fit the response. Stochastic 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.

Despite the important interest that kriging has raised both in the simulation and optimization community, the analysis of the convergence properties of kriging based algorithms represents an active field of research. Results on the asymptotic behavior of the EGO algorithm are provided in Locatelli (1997), and, more recently, Bull (2011) provided the analysis of the convergence rates of meta–model based search in a more general framework including the EGO.

The authors have previously worked in the stochastic settings by proposing the Two Stage Sequential Search Optimization algorithm (TSSO) (Quan et al. 2013) and, lately, they have been further improving the approach resulting in the extended–TSSO (eTSSO) (Pedrielli et al. 2015). This paper provides a theoretical understanding of the eTSSO algorithm, by taking a different perspective on kriging–based simulation–optimization. Specifically, we look at eTSSO as a stochastic recursion algorithm having EGO as deterministic counterpart. In particular, the parallelism between kriging–based search and stochastic recursion as well as the relationship between eTSSO and EGO are investigated to adapt the framework in Pasupathy et al. (2015) for the convergence analysis of kriging–based optimization algorithms.

The remainder of the paper is structured as follows: section 2 gives the main notation and terminology to support the preliminaries provided in section 3 where both the deterministic as well as the stochastic kriging–based approaches are presented. Section 4 presents the proposed adaptation of the framework in Pasupathy et al. (2015) to the case of stochastic–kriging based simulation–optimization. Section 5 offers some numerical evidence of the asymptotic convergence and convergence rate of the studied stochastic algorithm, while section 6 concludes the paper.

2 NOTATION AND TERMINOLOGY

In this section, we present some of the main definitions that will be used throughout the work.
We will refer to x as a vector of real numbers in a generic d–dimensional space $\mathbb{X} \subseteq \Re^d$. e_i is defined in $\Re^d$ and it denotes a unit vector whose i–th component is 1 and any other component is 0.

For a sequence of random variables, we say $\{X_n\} \xrightarrow{wp1} x$ to mean that the stochastic sequence $\{X_n\}$ converges to x with probability 1.
For a sequence of real numbers $\{a_n\}$, we say that $a_n = o(1)$ if $lim_{n\to\infty} a_n = 0$, and $a_n = O(1)$ if $\exists c \in (0,\infty)$ with $|a_n| < c$ for large enough n. Finally, we say that $a_n = \Theta(1)$ if $0 < \liminf a_n \leq \limsup a_n < \infty$.
We also adopt the following definitions to characterize the convergence behaviour of the analysed algorithms.

Definition 1 (Linear convergence) $\left\{x_k^\theta\right\}$ exhibits a *linear*(ℓ) convergence to x^* if $\limsup_{k\to\infty} \frac{||x_{k+1}-x^*||}{||x_k-x^*||} = \ell \in (0,1)$

The following definition characterizes the control of the sample size sequence we created for the stochastic algorithm eTSSO.

Definition 2 (Geometric growth of a sequence) A sequence $\{m_k\}$ exhibits geometric growth if $m_{k+1} = c \cdot m_k$, $k = 1,2,\ldots$ for some $c \in (1,\infty)$.

3 PRELIMINARIES

We consider a single objective minimization problem defined over a compact set $\mathbb{X}$. The deterministic d–dimensional objective function $f : x \in \mathbb{X} \subset \Re^d \to f(x) \in \Re$ is here observed running simulation experiment(s) at point x. Our goal is to find a global minimum of $f : \mathbb{X} \to \Re$, where $\mathbb{X} \subseteq \Re^d$ solving:

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

In order to find a solution to P, we use a search procedure based on a meta–model of the response, whose parameters are updated, as the search progresses, based on the simulation results. The model guides the search predicting the function values at points where no simulation has been conducted yet.

3.1 Deterministic Problem

We start with the case in which the simulation response is obtained without noise, i.e., in the case of deterministic optimization. In particular, suppose we wish to minimize an unknown function f, choosing design points x_k and the estimated minima x_k^*. We will refer to π as the statistical model estimating the behavior of the unknown function f. This model can be interpreted as our belief about the unknown function (Bull 2011). In this paper, we refer to the Efficient Global Optimization (EGO) algorithm, which, iteratively maximizing the *expected improvement*, generates a sequence of random design points $\{x_k\}$ and estimated best solutions $\{x_k^*\}$ taking value over the compact space $\mathbb{X}$. At iteration k, the filtration $\mathscr{F}_k$, defined as the sigma algebra $\sigma(x_i, f(x_i) : i \leq k)$, represents the set of available information. According to the EGO, given $\mathscr{F}_k$, we will choose as estimated best solutions at iteration k, x_k^*, the point, among those already sampled in the set $\mathbb{S} \subseteq \mathbb{X}$, having the best function value up to that iteration. The next point to sample is selected in order to maximize the following *Expected Improvement* function (Jones et al. 1998):

$$EI_{\pi_k}(x, \mathscr{F}_k) := \mathbb{E}_\pi \left[f(x_k^*) - \hat{f}(x) \,|\mathscr{F}_k \right] \tag{1}$$

where, as already stated, x_k^* represents the sampled point with the minimum associated function value $f\left(x_k^*\right)$ up to iteration k, and $\hat{f}(x)$ is the predicted function value at the non sampled point x. Typically, the function f is modeled as a stationary Gaussian process and we consider the values $f(x)$ of the non–sampled points x to be jointly Gaussian with mean and covariance parametrized through a constant τ and a d–dimensional vector θ, namely:

$$\mathbb{E}_\pi [f(x)] = \mu, \;\; Cov_\pi [f(x), f(y)] = \tau^2 K_\theta (x - y) \tag{2}$$

Having chosen a certain statistical model for f, each point $x \in \mathbb{X} \setminus \mathbb{S}$ is associated with the following predictor:

$$f(x) \,|(f(x_i)_{i \leq k}) \sim \mathscr{N}\left(\hat{f}(x;\theta), s_k^2(x;\theta)\right)$$

where:

$$\hat{\mu}_k(\theta) := \frac{1^T V^{-1} \mathbf{f}}{1^T V^{-1} 1} \tag{3}$$

$$\hat{f}_k(x, \theta) := \hat{\mu}_k(\theta) + v^T V^{-1} (\mathbf{f} - \hat{\mu}_k 1) \tag{4}$$

and

$$s_k^2(x, \theta) := \tau^2 \left(1 - v^T V^{-1} v + \frac{\left(1 - 1^T V^{-1} v\right)^2}{1^T V^{-1} 1} \right) \tag{5}$$

where, 1 is a vector having all elements equal to 1, $V = (K_\theta(x_i - x_j))_{i=1}^k$ is the spatial variance–covariance matrix and $v = (K_\theta(x_{k+1} - x_i))_{i=1}^k$ represents the correlation vector. Following Yin et al. (2011), we use a Gaussian kernel in equation (2), :

$$K_\theta(x_i - x_j) := \prod_{l=1}^{d} exp\left(-\theta_l \left(x_{il} - x_{jl}\right)^2\right) \tag{6}$$

Under assumptions 1–4 page 2883 in Bull (2011), which are satisfied in the present context, the author proves convergence rates for EGO. In particular, the author uses the Reproducing Kernel Hilbert Space $\mathscr{H}(X)$ of functions over the space $\mathbb{X}$ constructed from the kernel K and establishes the convergence rates of the loss function $L_k(EI_{\pi_k}, \mathscr{H}_\theta(X), R) := \sup_{||f||_{\mathscr{H}_\theta(X)} \leq R} \mathbb{E}_\pi \left[f\left(x_k^*\right) - \min \hat{f} \right]$ over the ball of radius R, B_R, in $\mathscr{H}(\mathbb{X})$ after k steps as (Theorem 2, page 2887, (Bull 2011)):

$$L_k(EI_{\pi_k}, \mathscr{H}_\theta(X), R) := \sup_{||f||_{\mathscr{H}_\theta(X)} \leq R} \mathbb{E}_\pi \left[f(x_k^*) - \hat{f}(x) \,|\mathscr{F}_k \right] = O\left(k^{-1/d}\right). \tag{7}$$

Under this result, considering the definitions in section 2, EGO exhibits *linear* convergence rates. In particular the rate is $\lim_{k\to\infty} \frac{||x_{k+1}-x^*||}{||x_k-x^*||} = O\left(\left(1-\frac{1}{k}\right)^{1/d}\right)$.

3.2 Stochastic Problem

In the stochastic context, a simulation is only able to return a point estimate of $f(x_i)$ for each replication run at the point $x_i \in \mathbb{X} \subseteq \Re^d$ and not its true value as in the deterministic case presented in section 3.1. In this setting, the authors refer to their recently proposed eTSSO (Pedrielli et al. 2015), a two–stage algorithm which uses the first stage to select a new point and subsequently assesses the number of replications required to re–evaluate the model parameters. The Modified Nugget Effect Kriging (MNEK) model (Yin et al. 2011) is used to estimate the function values at the non–sampled locations $x \in \mathbb{X} \notin \mathbb{S}$. Subsequently, the point x_k is added to the set $\mathbb{S}$ if it maximizes the *modified expected improvement* function $EI_{\tilde{\pi}_k}(x_i, \mathscr{F}_k)$, the stochastic version of (1).

Again, we assume that $f(x_i)$ are realizations of a random process and a statistical model $\tilde{\pi}_k$ represents the stochastic counterpart of π in section 3.1, namely:

$$f(x) = Z(x) + \xi(x), \;\; x \in \mathbb{X} \tag{8}$$

where Z describes the mean and ξ describes the random noise process. As in the deterministic case, we further model $Z(x)$ 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, $Z(x)$ is a $GP\left(\mu(x), \tau^2 V\right)$.

As already mentioned in section 3.1, 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 (equation (6)). 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 x (i.e., the heteroscedastic case is considered). With independent sampling (i.e., no CRN), V_ξ is diagonal, and equation (8) reduces to the independent sampling noise model (Yin et al. 2011). The general form of equation (8) is similar to the form proposed in Ankenman et al. (2010).

As shown in Yin et al. (2011), the predictor for (8) at the point x, given k points have been already sampled, is:

$$\hat{f}(W_k, x, \theta) = \sum_{i=1}^{k} \left(v^T \left(V_\xi + V\right)^{-1} e_i + 1^T \left(V_\xi + V\right)^{-1} \frac{\left[1 - 1^T \left(V_\xi + V\right)^{-1} v\right]^T}{1^T \left(V_\xi + V\right)^{-1} 1} e_i \right) \bar{f}_i \tag{9}$$

where, W_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, $v(W_k, x, \theta)^T = \left(e^{-\theta \cdot d_{x,x_1}^2} \cdots e^{-\theta \cdot d_{x,x_k}^2}\right)$, 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. The optimal MSE results (Yin et al. 2011):

$$s_k^2(W_k, x, \theta) := c_0 + \tau^2 \left(1 - \left[v + 1 \frac{\left(1 - 1^T V'^{-1} v\right)}{1^T V'^{-1} 1} \right]^T V'^{-1} v + \frac{\left(1 - 1^T V'^{-1} v\right)}{1^T V'^{-1} 1} \right) \tag{10}$$

where $V' = V + V_\xi$, and c_0 is the nugget effect value which usually can be estimated from the sample variance as $\hat{c}_0 = \hat{\sigma}^2 / W_{i,k}$, where $W_{i,k}$ represents the number of simulation replications allocated to point

Algorithm 1: eTSSO Algorithm

Initialization: Choose T, r_{min}, $\mathbb{S}: |\mathbb{S}| = N_0$, $m_{i0} = r_{min}$ and $W_0 = N_0 \cdot m_0$;
for $i = 1 \ldots, N_0$ **do**
  Run m_{i0} replications and return $\bar{f}_0(x_i), \hat{\sigma}^2_{\xi,0}(x_i)$;
end
Fit the MNEK model to the set of sample means and apply cross–validation to verify it.
Update $T = T - W_0$, $k = 1$.
while $W_{k-1} \leq T$ **do**
  Select $x_k \in \arg\max_{x \in \mathbb{X} \notin \mathbb{S}} EI_{\tilde{\pi}_k}(x)$, $\mathbb{S} \leftarrow \mathbb{S} \cup x_k$;
  Run r_{min} simulations to obtain $\left(\bar{f}_k(x_k), \hat{\sigma}^2_{\xi,k}(x_k)\right)$;
  Use OCBA to determine $\hat{\sigma}^2_{\xi,k}$ and update the budget m_k according to (12);
  $W_k = W_{k-1} + m_k$;
  if $(m_k - r_{min}) > 0$ **then**
    Apply equations (13)–(14) to allocate m_{ik} to the sampled points $i \in \mathbb{S}$;
    Run the simulation experiments according to the budget;
    Fit the kriging model according to the updated information;
    $k = k + 1$;
  end
end

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. In the rest of the paper, we will avoid the notation $(W_k, \cdot, \cdot)$ when it is clear we refer to the stochastic model and s^2_k will be referred to as *extrinsic variance*.
Algorithm 1 summarizes the steps of the eTSSO procedure. In its first stage, eTSSO, computes the function $EI_{\tilde{\pi}_k}(x_i, \mathscr{F}_k)$ over the set of non–sampled points $\mathbb{X} \setminus \mathbb{S}$ (Pedrielli et al. 2015):

$$x_k \in \arg\max_{x \in X \notin \mathbb{S}} EI_{\tilde{\pi}_k}(x_i, \mathscr{F}_k) := \mathbb{E}_{\tilde{\pi}_k}\left[\max\left\{\bar{f}(x_k^*) - \hat{f}(x), 0\right\}\right] \tag{11}$$

here, $\bar{f}(x_k^*)$ represents the response at the sampled points with the lowest *average* function value $\bar{f}$ up to iteration k, and $\hat{f}(x)$ is a random variable with the mean given by kriging mean function and the variance given by the spatial prediction uncertainty $s^2_k(x, \theta)$. The point x_k is sampled with r_{min} replications (with r_{min} being an input parameter), and added to the set of already sampled points $\mathbb{S}$.

At the second stage, eTSSO uses the Optimal Computing Budget Allocation (OCBA) technique (Chen et al. 2000) to assign the available simulation replications. Specifically, the sequence of simulation replications at each iteration is dynamically updated according to the following stochastic rule:

$$m_k = \lfloor m_{k-1}\left(1 + \frac{\hat{\sigma}^2_{\xi,k}}{\hat{\sigma}^2_{\xi,k} + s^2_k(x_k)}\right)\rfloor \tag{12}$$

where, $\hat{\sigma}^2_{\xi,k}$ refers to the estimated variance of the sampled point receiving more budget according to OCBA, whereas $s^2_k(x_k)$, formulated in (10), refers the location maximizing the function $EI_{\tilde{\pi}_k}(x_i, \mathscr{F}_k)$. The total budget used up to iteration k is $W_k = \sum_{k=0}^{k} m_k$. The budget at the first iteration is set to the minimum number of replications to sample a new point, i.e., $m_0 = r_{min}$. Since N_0 points are chosen for the initial MNEK model fit, $W_0 = N_0 \cdot m_0$ and the remaining budget is $T \leftarrow T - W_0$. At the generic k–th iteration the

budget m_k is allocated to the points in $\mathbb{S}$ using OCBA. In particular, equation (13) determines the relative allocation between non–best designs and equation (14) is used to derive the relative allocation between the best design and non best designs. If we refer to m_{ik} as the budget allocated to point i at iteration k, we have (Chen et al. 2000):

$$m_{ik}/m_{jk} = \left(\frac{\hat{\sigma}_{\xi,k}(x_i)/\delta_{b,i}}{\hat{\sigma}_{\xi,k}(x_j)/\delta_{b,j}} \right)^2, \tag{13}$$

$$m_{bk} = \hat{\sigma}_{\xi,k}(x_b) \sqrt{\sum_{x\in\mathbb{S}:x\neq x_b} \frac{m_{ik}^2}{\hat{\sigma}_{\xi,k}^2(x_i)}}, \tag{14}$$

At iteration k, x_b is the design point with the lowest sample mean, m_{bk} is the related number of replications and $\hat{\sigma}_{\xi,k}(x_b)$ is the related sample standard deviation; m_{ik} is the number of replications performed at location i and $\hat{\sigma}_{\xi,k}(x_i)$ is the estimated standard deviation at that point. $\delta_{b,i}$ is the difference between sample mean at point i and the lowest sample mean.

4 MAIN RESULTS

We can interpret the EGO in Jones et al. (1998) adopting the perspective of stochastic recursion algorithms. In particular, at the k–th iteration, manipulating the definition in (1), we have:

$$x_{k+1} = x_k^* + dist\left(x_k^*, \arg\max_{x\in\mathbb{X}\setminus\mathbb{S}} EI_{\pi_k}(x)\right) \tag{15}$$

where the function $dist(\cdot,\cdot)$ is the vector of the distances between the components of x_k^* and the candidate point x_{k+1}.

According to algorithm 1, we can interpret eTSSO as the stochastic counterpart of (15) and formulate the related iteration as:

$$X_{k+1} = X_k^* + dist\left(X_k^*, \arg\max_{x\in X\setminus\mathbb{S}} EI_{\tilde{\pi}_k}(x)\right) \tag{16}$$

where, $\tilde{\pi}_k$ refers to the model in (8) which replaces π in (2).

The main difference between (15) and the typical recursion in a deterministic search algorithm, resides in the filtration $\mathscr{F}_k$. According to the traditional update:

$$x_{k+1} = x_k + h(x_k) \tag{17}$$

whereas, in (15), the entire *sequence* of the visited points is considered at each step $k = 1,\ldots$:

$$x_{k+1} = g\left(\{x_i\}_{i=1}^k\right) + h\left(\{x_i\}_{i=1}^k\right) \tag{18}$$

here, $g\left(\{x_i\}_{i=1}^k\right) := x_k^* \in \arg\min_{x\in\mathbb{S}} f(x)$ and $h\left(\{x_i\}_{i=1}^k\right) := dist\left(x_k^*, \arg\max_{x\in\mathbb{X}\setminus\mathbb{S}} EI_{\pi_k}\right)$.
In eTSSO, we estimate the components in the iteration (18) as it follows:

$$X_{k+1} = G\left(W_k, \{X_i\}_{i=1}^k\right) + H\left(W_k, \{X_i\}_{i=1}^k\right) \tag{19}$$

where, $G\left(W_k, \{X_i\}_{i=1}^k\right) := X_k^* \in \arg\min_{x\in\mathbb{S}} \bar{f}(x)$, whereas $H\left(W_k, \{X_i\}_{i=1}^k\right) :=$
$:= dist\left(X_k^*, \arg\max_{x\in X\setminus\mathbb{S}} EI_{\tilde{\pi}_k}\right)$. W_k represents the total simulation budget used up to iteration k.

In the following, we will exploit the proof framework proposed in Pasupathy et al. (2015) for stochastic recursion, in order to study the eTSSO algorithm having the EGO as *Deterministic Analogue* (Jones et al. 1998). The following assumptions are required in the scope of showing our results.

Assumption 1 $\mathbb{X}$ is a compact space.

Assumption 2 Each dimension in the space is defined between $[0,1]$.

Assumption 2 simply requires normalization of the function domain $\mathbb{X}$.

Assumption 3 The initial sample size is such to produce an initial fit of the MNEK model satisfying cross-validation criteria.

Assumption 4 The Gaussian correlation function is adopted to model the spatial variance covariance matrix V.

Assumption 4 is a sufficient condition for the existence of the derivative processes and it ensures that the various variance–covariance matrices are positive definite, i.e., non singular. These will be used in Lemma 1, which characterizes the expected improvement function in (1).

Assumption 5 The parameters τ and θ of the MNEK model are assumed known.

Assumption 6 The number of replications m_k assigned at each iteration satisfies $m_k \geq m_{k-1},\ \forall k = 1,2,\ldots$ and $m_k \to \infty$ as $k \to \infty$. Moreover, for any $\varepsilon > 0$ there exists a $\delta_\varepsilon \in (0,1)$ and a $\bar{k}_\varepsilon > 0$ such that $\psi^{2k}\mathscr{L}(m_{k-1},\varepsilon) \leq (\delta_\varepsilon)^k,\ \forall k \geq \bar{k}_\varepsilon$, where $\mathscr{L}(\cdot,\cdot)$ is strictly decreasing in m_{k-1} and non–increasing in ε.

Assumption 7 The true function f to be optimized over the compact space $\mathbb{X}$ has a unique minimum x^*.

The following lemma characterizes the function $h(x)$ in the EGO iterates.

Lemma 1 There exists $\kappa \in \Re$ such that, for any $(x,y) \in \mathbb{X}$, $||h(x)-h(y)|| \leq \kappa||x-y||$.

Proof. To guarantee the result, we have to prove that, given two sequences $\{x_i\}_{i=1}^k$ and $\{\tilde{x}_i\}_{i=1}^k$ corresponding to the filtration $\mathscr{F}_k$ and $\tilde{\mathscr{F}}_k$, respectively, which are *close* in terms of euclidean distance between the related points, the resulting processes $h(x)|f(x_i)_{i\leq k}$ and $h(x)|f(\tilde{x}_i)_{i\leq k}$ are also close. From the definition in equation (15), these sequences are generated according to the expected improvement. As a result, requiring close sequences corresponds to guaranteeing the function EI_{π_k} is Lipschitz continuous.
We can formulate the expected improvement as (Locatelli 1997):

$$EI_{\pi_k}(x_i;\mathscr{F}_k) = s_k(x_i,\theta)\phi\left(\frac{f(x_k^*)-\hat{\mu}_k(\theta)}{s_k(x_i,\theta)}\right) - (f(x_k^*)-\hat{\mu}_k(\theta))\left(1-\Phi\left(\frac{f(x_k^*)-\hat{\mu}_k(\theta)}{s_k(x_i,\theta)}\right)\right), \quad (20)$$

where ϕ and Φ represent the *pdf* and *cdf* of the normal distribution, respectively. Such a function is Lipschitz continuous in case $\hat{\mu}_k(\theta)$ and $s_k(x_i,\theta)$ have the form of equations (3) and (5), respectively.
To show Lipschitz continuity we have to guarantee that $\frac{dEI_{\pi_k}(x_i;\mathscr{F}_k)}{dx} < \infty,\ \forall x \in \mathbb{X}$. Since we consider Gaussian processes, the components $\phi\left(\frac{f(x_k^*)-\hat{\mu}_k(\theta)}{s_k(x_i,\theta)}\right)$ and $\Phi\left(\frac{f(x_k^*)_k-\hat{\mu}_k(\theta)}{s_k(x_i,\theta)}\right)$ are the *pdf* and *cdf* of a normal distribution with finite mean and variance $0 < \hat{\mu}_k(\theta), s_k^2(x_i,\theta) < \infty$, and finite derivative. More attention needs to be paid to the derivatives of $\hat{\mu}_k(\theta)$ and $s_k(x_i,\theta)$. We rewrite in explicit form equation (5):

$$s_k(x_i,\theta) = \tau\left(1-\left(\sum_{h=1}^k\sum_{g=1}^k e^{-\sum_{j=1}^d \theta_j\left(x_{ij}-x_{hj}\right)^2} e^{-\sum_{j=1}^d \theta_j\left(x_{gj}-x_{hj}\right)^2} r_{hg}^{-1}\right)\right)^{1/2} \quad (21)$$

It is apparent that under assumption 4, equation (21) is infinitely differentiable with respect to x_i. The finiteness of the derivative, however, depends on the values of $r_{hg}^{-1},\ \forall(h,g)$ representing the terms of the matrix V (h–th row, g–th column). Hence, the derivative will exist finite as long as the matrix V is non–singular, i.e., if assumption 4 holds. The same can be proven for the mean component. We then conclude function $EI_{\pi_k}(x_i;\mathscr{F}_k)$ is Lipschitz continuous. □

Lemma 2 The EGO algorithm which uses the Gaussian correlation form satisfies $\lim_{k\to\infty} x_k = x^*$. Moreover, the convergence is uniform in $\{x_i\}_{i=1}^{N_0}$, i.e., it is independent from the specific initial set of points chosen to generate the first model estimate.

Proof. From (21), we observe it is possible to bound the term $s_k(x_i,\theta)$ from above as follows:

$$s_k(x_i,\theta) = \tau\left(1 - \left(\sum_{h=1}^{k}\sum_{g=1}^{k} e^{-\sum_{l=1}^{d}\theta_j(x_{il}-x_{hl})^2} e^{-\sum_{l=1}^{d}\theta_j\left(x_{gl}-x_{hl}\right)^2} r_{hg}^{-1}\right)\right)^{1/2} \leq \tau \tag{22}$$

From Lemma 1, we have $EI_{\pi_k}(x_i;\mathscr{F}_k)$ is Lipschitz continuous. Hence, $EI_{\pi_k}(x_i;\mathscr{F}_k)$ is finite over the design space and, given two points, $|EI_{\pi_k}(x_i;\mathscr{F}_k) - EI_{\pi_k}(x_j;\mathscr{F}_k)|$ is finite as well. We can then consider the following:

$$EI_{\pi_k}(x_i;\mathscr{F}_k) \leq s_k(x_i,\theta)\phi\left(\frac{f\left(x_k^*\right) - \hat{\mu}_k(\theta)}{s_k(x_i,\theta)}\right) \leq s_k(x_i,\theta)/\sqrt{2\pi} \leq \tau/\sqrt{2\pi}, \tag{23}$$

where the last two inequalities are derived from the normality property and equation (22), respectively. According to (22) and (23), we can use the results in Locatelli (1997). Indeed, we note that in (22), the variance in the new point x_i is a weighted sum of the estimates at the already sampled points, and weights are a function of the euclidean distance between the candidate point and the sampled points. This is consistent with the form recognized in Locatelli (1997), which we extended to the d–dimensional case due to assumptions 2–6 and by replacing linear with euclidean distances. As a result of (23) and Lemma 1, we can apply Lemma 1 in Locatelli (1997) (page 60) using the Lipschitz constant $c = \tau$, leading to the following result ((Locatelli 1997), page 61):

$$\lim_{k\to\infty}\max_{i,j\in\mathbb{S}} ||x_i, x_j|| = 0. \tag{24}$$

Hence, the set of points at which the function is observed if the algorithm is never stopped, is dense in $\mathbb{X}$, proving convergence of the algorithm.

Concerning the uniform convergence result, there exists a number of initial points N_0 such that the convergence is guaranteed independently from the specific initial set (Bull 2011), and, under Assumption 3, we satisfy this condition. □

Now, we are ready to characterize the behaviour of eTSSO as the stochastic counterpart of EGO.

Lemma 3 As the number of iterations $k \to \infty$, under assumptions 6–7, the MNEK model $\tilde{\pi}_k$ approaches its deterministic counterpart π.

Proof. If assumption 6 holds, let us consider $\mathscr{L} = Tr\left(\sigma_\xi^2 V_\xi\right)$, where $Tr(\cdot)$ is the trace of a matrix, and let $\delta_\varepsilon = 1/\tau^2$. The function $Tr\left(\sigma_\xi^2 V_\xi\right)$ satisfies the properties required to $\mathscr{L}$: being an error function, it is strictly decreasing in W_k. The assumption is stating that there exist a finite number of iterations such that $\mathscr{L} \leq \left(\delta_\varepsilon/\psi^2\right)^k$. This means that at each iteration, the algorithm produces estimates of $\mathscr{L}$ which are decreasing with $\delta_\varepsilon/\psi^2$. To show this holds, let us rewrite the covariance matrix:

$$V' = V + V_\xi = \begin{pmatrix} 1 & e^{\left(-\theta\cdot d_{12}^2\right)} & \dots & e^{\left(-\theta\cdot d_{1k}^2\right)} \\ e^{\left(-\theta\cdot d_{21}^2\right)} & 1 & \dots & e^{\left(-\theta\cdot d_{2k}^2\right)} \\ \vdots & \vdots & \vdots & \vdots \\ e^{\left(-\theta\cdot d_{k1}^2\right)} & \dots & \dots & 1 \end{pmatrix} + \begin{pmatrix} \frac{\sigma_\xi^2(x_1)}{W_{1,k}\tau^2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \frac{\sigma_\xi^2(x_k)}{W_{k,k}\tau^2} \end{pmatrix} \tag{25}$$

Here, d_{ij} represents the euclidean distance between two points (i, j), $W_{i,k}$ represents the number of replications performed at location i up to iteration k according to the eTSSO budget allocation scheme. Having assumed knowledge on the kriging variance (assumption 5), the diagonal elements of the matrix V_ξ are bounded by decreasing values of $\delta_\varepsilon/\psi^2$, where $\delta_\varepsilon = \frac{1}{\left(\max_{x_j\in\mathbb{S}} W_{j,k}\right)}$ and $\psi^2 = \tau^2$ ($\mathbb{S}$ refers to the set of sampled points). Apparently, $\delta_\varepsilon \in (0,1)$; moreover, $\left(\left(\max_{x_j\in\mathbb{S}} W_{j,k}\right)\tau^2\right)^{-1} \to 0$ as $k\to\infty$, since $\max_{x_j\in\mathbb{S}} W_{j,k} \to \infty\ \forall k$. Hence, $V_\xi \xrightarrow{wp1} 0$. □

Lemma 4 As the number of iterations $k\to\infty$, $\hat{f}(x|\mathscr{F}_k) \sim N\left(\mu_{|\mathscr{F}_k}(x), \sigma^2_{|\mathscr{F}_k}(x)\right)$.

Proof. As the number of iterations $k\to\infty$ we can apply Lemma 3 and the result in (Stein 1999) (Appendix A). $\hat{f}(x|\mathscr{F}_k)$ has normal distribution with mean and variance:

$$\hat{f}(W_k,x,\theta) \xrightarrow{k\to\infty} \left(v^T V^{-1} e_i + 1^T V^{-1}\frac{\left[1-1^T V^{-1} v\right]^T}{1^T V^{-1} 1} e_i\right)\bar{f}(x) \tag{26}$$

$$s_k^2(W_k,x,\theta) \xrightarrow{k\to\infty} \tau^2\left(1-\left[v+1\frac{\left(1-1^T V^{-1}v\right)}{1^T V^{-1}1}\right]^T V^{-1}v + \frac{\left(1-1^T V^{-1}v\right)}{1^T V^{-1}1}\right) \tag{27}$$

corresponding to the deterministic predictor and variance in (Yin et al. 2011) equivalent to (3)–(5). □

At this point, we want to characterize the stochastic iteration in (16). In order to do so, we study $G(\cdot,\cdot)$ as well as $H(\cdot,\cdot)$. If we consider $G(\cdot,\cdot)$, at each iteration, the best point is the one having the best sampled mean and not the best true value, which cannot be evaluated, and the following property holds.

Property 1 (Convergence of $G\left(W_k, X_k^*\right)$) For any $\delta>0$, and with $k\to\infty$, $sup_{x\in\mathbb{X}} Pr\left\{\left(G(W_k,X_k^*) - g(x_k^*)\right) > \delta\right\} = O\left(W_k\left(X_k^*\right)^{-1/2}\right)$.

Proof. Here, $W_k\left(X_k^*\right)$ is the total budget allocated to point X_k^*. Due to the geometric growth of the replication in equation (12) $W_k\to\infty$ when $k\to\infty$. Due to Lemma 3, we simply exploit the Law of Large numbers to argue that the sample average will converge to the true function value provided that the sampling effort goes to infinity. In fact, we know that x^* will be sampled *infinitely often* due to the fact that OCBA rule is used to allocate the budget to each point to sample at each iteration. Due to the results in (Chen et al. 2000), we know that the best point will be sampled infinitely often (Theorem 1, page 258). As a result, Property 1 holds. □

The next results focus on $H(W_k,x)$ in order to characterize the eTSSO efficiency and consistency.

Theorem 1 Let $k\to\infty$ *(i)* As $W_k\to\infty$, the estimator $H(W_k,x)$ satisfies, for any $\Delta>0$, that $sup_{x\in\mathbb{X}} Pr\{||H(W_k,x)-h(x)||>\Delta\} = O\left(W_k^{-2\alpha}\right)$. *(ii)* If the sample sizes $\{W_{i,k}\}$ satisfy $W_{i,k}\to\infty$, then $||H(W_k,x)-h(x)|| \xrightarrow{wp1} 0$.

Proof. Result (i) is a direct consequence of Lemma 3 and Lemma 4 when $\alpha=\frac{1}{2}$. The sequence $\{W_{i,k}\}$ satisfies $W_{i,k}\to\infty$ due to equation (12), and we can now simply recall the results in Lemma 2 to prove part (ii). □

At this point, we can exploit the results in Pasupathy et al. (2015) to study the efficiency of Algorithm 1 in section 3.2. Before presenting the main result, we note the following.

Property 2 (Characterization of $H(W_k,x)$) Let $k\to\infty$, then the estimator $H(W_k,x)$ satisfies $sup_{x\in\mathbb{X}}\mathbb{E}(H(W_k,x)-h(x)) = \Theta\left(\left(\tau^2\cdot\min(W_{i,k})\right)^{-1}\right)$.

Proof. As already stated, $W_k \to \infty$ when $k \to \infty$. Under assumption 6 and the result in Lemma 1 $EI_{\tilde{\pi}_k}$ is lower semi–continuous. Lemma 1 guarantees the function is finite over the space $\mathbb{X}$. We can use the results in Attouch (1984) to say that $EI_{\tilde{\pi}_k}$ epi–converges to EI_{π} when $W_k \to \infty$. Concerning the optimal location, Theorem 1, page 258 in (Chen et al. 2000) guarantees infinite budget will be allocated to the best point. Then, given epi–convergence, we can simply apply Theorem 3.4 in Robinson (1996), to say that the sequence $EI_{\tilde{\pi}_k} \to EI_{\pi_k}$ and the sequence of selected points $X_k \in \arg\max EI_{\tilde{\pi}_k} \to x_k$. This convergence is determined by the model and the variance covariance matrix in equation (25). Hence, as a consequence of Lemma 3 and Theorem 1, the property holds. □

Theorem 2 (Convergence rate of Algorithm 1) Let us define $c_k := 1 + \frac{\hat{\sigma}^2_{\xi,k}}{\hat{\sigma}^2_{\xi,k} + s^2_k(x_k)}$ and $\ell = \left(1 - \frac{1}{k}\right)^{1/d}$. Given that EGO exhibits linear convergence, for any $\varepsilon > 0$ satisfying $\ell + \varepsilon < 1$ and as $k \to \infty$, the following holds for $E_k = ||X_k - x^*||$:

$$\text{if } c_k \in \left(1, \ell^{-2}\right), \qquad E_k = O(W_k^{-1/2}) \tag{28}$$

$$\text{if } c_k \geq \ell^{-2}, \quad E_k = O\left(\left(c_k^{-1/2}(\ell+\varepsilon)^{-1}\right)^{-k} W_k^{-1/2}\right) \tag{29}$$

Proof. According to Bull (2011) the error decay rate of EGO, under assumptions 1–4 page 2883, is $\ell = \left(1 - \frac{1}{k}\right)^{1/d}$. According to theorem 1, we have that $sup_{x \in \mathbb{X}} Pr\{||H(W_k, x) - h(x)|| > \Delta\} = O\left(W_k^{-2\alpha}\right)$ with $\alpha = 1/2$. From (12), we observe that the coefficient c_k for the geometric increase of the budget at each algorithm iteration satisfies $c_k \leq 2$, also $c_k \xrightarrow{k=\infty} 1$. Since the budget increase is stochastic, we need to consider both the case $c_k \in \left(1, \ell^{-2}\right)$ and $c_k \geq \ell^{-2}$ in (Pasupathy et al. 2015) and, considering the definitions of ℓ and c_k, we use Theorem 6.5 page 17 in Pasupathy et al. (2015) to prove the result. □

It has to be noted that, in assumption 5, we stated that the parameters θ and τ^2 of the MNEK model are known in advance. However, convergence is affected by the quality of the estimators (Kleijnen et al. 2012). If MLEs are adopted, in deterministic settings, Bull (2011) proves convergence is guaranteed assuming to be able to produce a bounded estimate of $\hat{\theta}$. Intuitively, when the bias is consistent, the optimal location should still be identified.

5 EMPIRICAL RESULTS

We empirically observed the convergence of eTSSO in the location and the function estimation, by evaluating the number of iterations to let the euclidean distances $|x - x^*|$ and $|y - y^*|$ converge to 0, respectively (where x^* is the reachable optimum in a finite grid (Kleijnen et al. 2012)). This analysis provides empirical evidence of eTSSO convergence even when the model parameters are sequentially estimated. We present a 1–d, a 2–d tetra–modal, and the 3–d Hartmann function. For the 1–d case, we used the function:

$$f(x) = (2x + 9.96)\cos(13x - 0.26) + \xi(x).$$

We applied a normal random noise having variance $\sigma^2_\xi = \delta \cdot (x_1)$, with $\delta = \{0.1, 1.0, 10.0\}$.

We noticed that the convergence rate is affected by the noise level. In particular, in the case $\delta = 0.1$, the location is identified after 168 simulation replications (corresponding to 6 iterations of the eTSSO algorithm), 7733 (corresponding to 13 iterations) in the case $\delta = 1.0$ and 32094 (corresponding to 15 iterations) for the case $\delta = 10.0$. It is also noteworthy how good performances are already reached for much lower budgets: in the low noise case, an error of 0.001 in the location is reached after 80 simulation replications (i.e., two algorithm iterations), for $\delta = 1.0$, 136 replications are required (i.e., four iterations) and in the case of $\delta = 10.0$, 1094 replications were required, corresponding to 10 iterations.
For the 2–d case, we considered the following tetra–modal:

$$f(x_1, x_2) = -5(1 - (2x_1 - 1)^2)(1 - (2x_2 - 1)^2)(4 + 2x_1 - 1)\left(0.05^{(2x_1-1)^2} - 0.05^{(2x_2-1)^2}\right)^2 + \xi(x).$$

Table 1: Parameters A_{ij} and P_{ij} of the Hartmann–3 function

A_{ij}			P_{ij}		
3	10	30	0.3689	0.117	0.2673
0.1	10	35	0.4699	0.4387	0.747
3	10	30	0.1091	0.8732	0.5547
0.1	10	35	0.03815	0.5743	0.8828

Table 2: Theoretical versus Empirical convergence rates.

	Average Empirical Convergence Rate			Average Theoretical Convergence Rate		
	$\delta = 0.1$	$\delta = 1.0$	$\delta = 10.0$	$\delta = 0.1$	$\delta = 1.0$	$\delta = 10.0$
1–d	0.891667	0.95378788	0.9619742	0.265934	0.120082	0.1253412
Tetra–Modal	0.842135	0.97017507	0.96720977	0.313284	0.269857	0.2251032
Hartmann–3	0.587	0.46614123	1.28771458	0.235702	0.229416	0.22369009

Both dimensions of the test function, x_1 and x_2, are scaled to $[0,1]$. The global minimum is located at $[0.85, 0.5]$ and has the value -7.098. We applied to the function a normal random noise with variance $\sigma^2_\xi = \delta \cdot (|x_1| + |x_2|)$ and we tried three noise scenarios, namely $\delta = \{0.1, 1.0, 10.0\}$.
In this case, the convergence is reached after 272 simulation replications (7 algorithm iterations) in the case $\delta = 0.1$, 1071 (17 iterations) replications are required for the case $\delta = 1.0$, while 115852 replications are required when $\delta = 10.0$ corresponding to 44 iterations of eTSSO. As in the previous case, however, reasonable result with an error in the order $10e(-3)$ are obtained already with a number of replications equal to 220 (i.e., two iterations), 352 (i.e., six iterations) and 5955 (i.e., 13 iterations) for the three noise levels $\delta = \{0.1, 1.0, 10.0\}$, respectively. For the 3–d case, we adopt the following Hartmann–3 function:

$$f(x_1, x_2, x_3) = -\sum_{i=1}^{4} \alpha_i \exp\left[-\sum_{j=1}^{3} A_{ij}(x_j - P_{ij})^2\right] + \xi(x).$$

Here, $0 \le x_i \le 1$ for $i = 1,2,3$; parameters $\alpha = (1.0, 1.2, 3.0, 3.2)$, and A_{ij} and Pij given in Table 1. The function has a global minimum at $x^* = [0.114614, 0.555649, 0.852547]$ with $f(x^*) = -3.86278$; the function has three additional local minima. We applied to the function a normal random noise with variance $\sigma^2_\xi = \delta \cdot (|x_1| + |x_2| + |x_3|)$ and we tried the following set of values for δ, $\{0.1, 1.0, 10.0\}$.
A similar behavior with respect to the previous cases was observed. Nevertheless, the increased dimension leads to a slower convergence rate in the optimum location. Specifically, in this case, the convergence of $|x - x^*|$ was reached only with budget 250000 (corresponding to 16 iterations). This has to be brought back also to the noise of the Hartmann–3 which is larger than in the previous cases. Table 2 shows the average convergence rate derived from the iterations of the algorithm where, as theoretical counterpart, we use $W_k^{-1/2}$. We considered $k = 1, \ldots, K^*$, being K^* the iteration at which the algorithm reaches convergence.

6 CONCLUSIONS

eTSSO is a kriging–based algorithm recently proposed by the authors extending the TSSO. Its efficiency builds upon a stochastic dynamic sampling rule which geometrically increases the budget to assign to each algorithm iteration. Inspired by the proof framework in Pasupathy et al. (2015), we propose to apply it to characterize consistency and efficiency of the eTSSO iterates. To do this, we exploit the properties of the EGO algorithm and interpret eTSSO as its stochastic counterpart. This paper is a first step towards a general approach for the performance analysis of meta–model based simulation–optimization algorithms. According to the main theorem 2, eTSSO is efficient only in case the coefficient for the geometric increase of the budget satisfies $c_k \in \left(1, \ell^{-2}\right)$. The stochastic nature of the budget allocation does not guarantee

this. Nevertheless, $c_k \in \left(1, \ell^{-2}\right)$, and this makes the eTSSO allocation closer to the efficient case and the empirical results sustain this argument. Further research will generalize the study to meta–model based stochastic algorithms and consider sequential parameters estimation.

ACKNOWLEDGMENTS

Research supported in part by the grant (R-SMI-2013-MA-11) funded by the Singapore Maritime Institute.

REFERENCES

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

Attouch, H. 1984. "Variational Convergence of Functions and Operators".

Bull, A. D. 2011. "Convergence rates of efficient global optimization algorithms". *The Journal of Machine Learning Research* 12:2879–2904.

Chen, C.-H., J. Lin, E. Yücesan, and S. E. Chick. 2000. "Simulation budget allocation for further enhancing the efficiency of ordinal optimization". *Discrete Event Dynamic Systems* 10 (3): 251–270.

Fu, M. C., C.-H. Chen, and L. Shi. 2008. "Some topics for simulation optimization". In *Proceedings of the 40th Conference on Winter Simulation*, 27–38. Winter Simulation Conference.

Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient global optimization of expensive black-box functions". *Journal of Global optimization* 13 (4): 455–492.

Kleijnen, J. P., W. van Beers, and I. Van Nieuwenhuyse. 2012. "Expected improvement in efficient global optimization through bootstrapped kriging". *Journal of global optimization* 54 (1): 59–73.

Locatelli, M. 1997. "Bayesian algorithms for one-dimensional global optimization". *Journal of Global Optimization* 10 (1): 57–76.

Pasupathy, R., P. Glynn, S. Ghosh, and F. Hashemi. 2015. "How much to sample in simulation–based stochastic recursions". *http://lebox.vt.edu/users/pasupath/pasupath.htm*.

Pedrielli, G., S. H. Ng, and C. Liu. 2015. "eTSSO: Adaptive Search Method for Stochastic Global Optimization Under Finite Budget". *Available at www.maritimestudies.nus.edu.sg/CMS %Research%20Highlight%20-%20August%202014.pdf*.

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.

Robinson, S. 1996. "Analysis of Sample–Path Optimization". *Mathematics of Operations Research* 21:513–528.

Stein, M. L. 1999. *Interpolation of spatial data: some theory for kriging*. Springer.

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.

AUTHOR BIOGRAPHIES

GIULIA PEDRIELLI is Research Fellow for the Centre for Maritime Studies at the National University of Singapore. Her research focuses on stochastic simulation-optimization in both single and multiple–objectives framework. Her email address is cmsgp@nus.edu.sg.

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.