AN ADAPTIVE RADIAL BASIS FUNCTION METHOD USING WEIGHTED IMPROVEMENT

Yibo Ji
Sujin Kim

Department of Industrial and Systems Engineering
National University of Singapore
119260, SINGAPORE

ABSTRACT

This paper introduces an adaptive Radial Basis Function (RBF) method using weighted improvement for the global optimization of black-box problems subject to box constraints. The proposed method applies rank-one update to efficiently build RBF models and derives a closed form for the leave-one-out cross validation (LOOCV) error of RBF models, allowing an adaptive choice of radial basis functions. In addition, we develop an estimated error bound, which share several desired properties with the kriging variance. This error estimate motivates us to design a novel sampling criterion called weighted improvement, capable of balancing between global search and local search with a tunable parameter. Computational results on 45 popular test problems indicate that the proposed algorithm outperforms several benchmark algorithms. Results also suggest that multiquadrics introduces lowest LOOCV error for small sample size while thin plate splines and inverse multiquadrics shows lower LOOCV error for large sample size.

1 INTRODUCTION

In this paper, we focus on finding the global minimum of a real-valued deterministic function $f(x)$, $x \in \mathscr{X}$, where $\mathscr{X}$ is a subset of $\mathbb{R}^d$ defined by box constraints and f is assumed to be lower-bounded, and continuous. We assume that $f(x)$ has no closed mathematical form and can only be evaluated through simulation.

One common approach to deal with such problems is to build a metamodel and use the minimizer of the metamodel as an approximate optimal solution (Booker et al. 1999). However, the error introduced by the metamodel approximation should be carefully examined since it may mislead the algorithm to a wrong solution wasting a given computational budget. In case where a kriging model is used as a metamodel, there is a well-established analysis on prediction error, and candidate solutions can be chosen by taking it into consideration. Efficient Global Optimization (EGO) by Jones et al. (1998) is one of the most well-known kriging-based global optimization algorithm, in which a new candidate point is determined by maximizing the expected improvement function. Jones (2001) provided in-depth reviews on sampling strategies with kriging models.

RBF models are also widely used in metamodel-based global optimization. They can be trained efficiently by solving a system of linear equations and are applicable in almost any dimensional problems since few restrictions are imposed on the location of sample points. The literature on RBF model-based optimization has two main branches. One employs a utility function defined for RBF metamodels. Assuming that the underlying objective function is smooth with a few local solutions, Gutmann (2001) defined a bumpiness measure for RBF models. Provided with a target value (an estimate of the optimal objective value), this method selects a new candidate point that minimizes the bumpiness of RBF models. Improved strategies of specifying the target value are presented in Björkman and Holmström (2000) and Regis and Shoemaker (2007a). The other branch of the literature focuses on the better utilization of objective function values predicted by RBF models and prediction uncertainties represented by the distance to the nearest evaluated point. The balance between exploration (maximize the prediction uncertainty) and exploitation

957

(minimize the predictive value) is achieved by applying different weighting scheme on these two pieces of information. Regis and Shoemaker (2005) exploited RBF metamodels by constraining the minimizer to be distant from evaluated points. Regis and Shoemaker (2007b) and Regis (2011) sequentially evaluated sample points with the maximum weighted score between the RBF metamodel prediction and the distance measure. Jakobsson et al. (2010) estimated the prediction uncertainty using the same distance measure and introduced a quality function defined by the total weighted uncertainty. Wendland (2005) reviewed an error bound for RBF models, but its dependence on a parameter related to the unknown true function f restricts its applications in metamodel-based optimization.

All these distance measures only provide a relative and crude estimate of the prediction uncertainty. A better error estimate may guide the algorithm to conduct global search more wisely. To our knowledge, the literature does not provide an easy-to-use and mathematical error estimate for RBF models, which allows RBF-based optimization algorithms to efficiently select basis functions. This paper makes the following contributions to the field of metamodel-based global optimization.

First, we propose a scheme to update RBF models using the Woodbury formula (Golub and Van Loan 1996), which allows us to efficiently inverts the interpolation matrix of RBF models whenever additional points are available. The same formula helps us obtain a closed form of the LOOCV error of RBF models. Based on this result, WIRBF can adaptively select radial basis functions throughout iterations. Second, inspired by the bumpiness measure developed by Gutmann (2001), we provide a new point-wise error estimate for RBF predictions using LOOCV. This error estimate can be efficiently computed and shares several properties with the variance function of kriging models. Based on this new error estimate and the smoothness measure of RBF models, we introduce the weighted improvement to balance the global search and local search with a tunable parameter. The WI at an unknown point considers the possible improvement introduced when assuming its true function value falls in the estimated error bound of the RBF prediction. For each possible improvement, the WI function assigns a weight derived based on the smoothness of the RBF model. Points with large prediction error and low prediction value show higher weighted improvement.

By incorporating all the above results, we propose an adaptive radial basis function method WIRBF. One feature of WIRBF is that it interrupts unpromising iterations and restarts with new sample points. However, it differs from existing multistart-based algorithms such as Regis and Shoemaker (2007b) in the sense that WIRBF requires new sample points to be distant from existing points. The new sample points are sequentially generated by maximizing the prediction error. We provide data profiles of WIRBF and four well-known metamodel-based global optimization algorithms on a set of 45 test problems from the literature.

This paper is organized as follows. Section 2 reviews radial basis function interpolation models. Section 3 lays out the proposed radial basis function method WIRBF and discusses its main components in details. Section 4 reports the data profiles of WIRBF and other existing algorithms on a set of test problems. Section 5 concludes this paper.

2 RADIAL BASIS FUNCTIONS MODELS

The standard RBF interpolation method, first discovered by Hardy (1971), approximates the unknown function with a linear combination of radially symmetric functions.

Definition 1 A standard RBF model interpolating n distinct evaluated points $S_n = \{(x_i, y_i = f(x_i))\}_{i=1}^n$ takes the form of $M_{S_n}(x) = \sum_{i=1}^n \alpha_i \psi(\|x - x_i\|)$, where $\psi(\cdot)$ is a real-valued function on $[0, \infty)$, and $\alpha = \{\alpha_i\}_{i=1}^n$ are determined by the interpolation conditions $M_{S_n}(x_i) = y_i$ for all $i = 1, \ldots, n$.

In this paper, an augmented RBF model with extra polynomial summands is also considered. This model has been extensively studied by Powell (1992) and Wendland (2005).

Definition 2 An augmented RBF model M_{S_n} interpolating n distinct evaluated points $S_n = \{(x_i, y_i)\}_{i=1}^n$ takes the form of

$$M_{S_n}(x) = \sum_{i=1}^n \alpha_i \psi(\|x - x_i\|) + \sum_{j=1}^{\widehat{\mathsf{m}}} \beta_j \pi_j(x), \tag{1}$$

where for all $i = 1, \ldots, n$ and $j = 1, \ldots, \widehat{\mathsf{m}}$, $\alpha_i \in \mathbb{R}$, $\beta_j \in \mathbb{R}$, and $\{\pi_j(x)\}_{j=1}^{\widehat{\mathsf{m}}}$ are bases of a $\widehat{\mathsf{m}} = \binom{\mathsf{m}-1+d}{d}$-dimensional linear space $\mathscr{P}_{\mathsf{m}}^d$ with polynomials in $\mathbb{R}^d$ of degree less than m.

Table 1 summarizes commonly-used radial basis functions, where $\theta > 0$ is a tunable shape parameter. In Section 3.2, we apply model selection techniques to determine a suitable ψ for a particular problem.

Table 1: Conditionally positive definite functions of different orders.

Name	$\psi(r)$	Order m_ψ
Multiquadrics (MQ)	$-\sqrt{r^2 + \theta}$	1
Thin plate splines (TPS)	$r^2 \ln(r)$	2
Cubic splines (CU)	r^3	2
Gaussian (GS)	$\exp(-r^2/\theta)$	0
Inverse multiquadrics (IMQ)	$(\theta + r^2)^{-1/2}$	0

For a particular ψ, we define matrix $\Psi_n \in \mathbb{R}^{n \times n}$ with its (i, j) element being $\psi(\|x_i - x_j\|)$ and matrix $\Pi \in \mathbb{R}^{n \times \widehat{\mathsf{m}}}$ whose (i, j) element is $\Pi_{ij} = \pi_j(x_i)$. Polynomials with high degrees are interesting but not attractive in the context of simulation optimization since they require more evaluated points for interpolation. In this paper, we use a linear tail ($\mathsf{m} = 2$, $\widehat{\mathsf{m}} = d + 1$) in (1), then the ith row of Π is $(1, x_i^\mathsf{T})$. The weight column vectors $\alpha = \{\alpha_i\}_{i=1}^n$, and $\beta = \{\beta_i\}_{i=1}^{\widehat{\mathsf{m}}}$ of model M_{S_n} are therefore obtained by solving the system $\begin{pmatrix} \Psi_n & \Pi \\ \Pi^\mathsf{T} & 0_{\widehat{\mathsf{m}} \times \widehat{\mathsf{m}}} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} y_n \\ 0_{\widehat{\mathsf{m}}} \end{pmatrix}$. Given distinct sample points S_n and the preceding choices of ψ, Powell (1992) proves that the interpolation matrix $\mathbf{C}_{\widehat{\mathsf{m}}+n} = \begin{pmatrix} \Psi_n & \Pi \\ \Pi^\mathsf{T} & 0_{\widehat{\mathsf{m}} \times \widehat{\mathsf{m}}} \end{pmatrix}$ is nonsingular if and only if the matrix Π has full column rank and the polynomial degree satisfies $\mathsf{m} \geq \mathsf{m}_\psi$ for each ψ. In addition, the interpolation matrix remains invertible while including additional distinct sample points.

Section 3.1 and 3.2 discuss augmented RBF models, while the proposed error estimates are derived based on standard RBF models for simplicity. A rigorous discussion on error estimates of augmented RBF models requires additional technical details. All results, however, can be easily applied to both types of RBF models.

3 THE ALGORITHM

Figure 1 depicts the general structure of the proposed algorithm, WIRBF, for the global optimization of black box problems. The full algorithmic description is presented in Algorithm 1.

WIRBF is a new multistart global optimization algorithm, which takes advantage of evaluated points according to a new strategy (see Step 3 and 4). Throughout the process, WIRBF records evaluated points with two sets: full information $S_n = \{(x, f(x)) : x \in \mathbf{X}_n\}$ and partial information $\widetilde{S}_n = \{(x, f(x)) : x \in \widetilde{\mathbf{X}}_n\}$, where $\mathbf{X}_n = \{x_i\}_{i=1}^n$ is the set of all evaluated sample points so far and $\widetilde{\mathbf{X}}_n$ is the set of evaluated points since the most recent *Restart*. It is worth noting that $\mathbf{X}_n = \bigcup_{i=1}^n \widetilde{\mathbf{X}}_i$ and $\widetilde{S}_n$ only records points evaluated between two consecutive restarts. Hence $\widetilde{S}_n = S_n$ until the first restart and $\widetilde{S}_n \subsetneq S_n$ after WIRBF restarts.

WIRBF starts by constructing a set $\mathbf{X}_{n_0} = \{x_i\}_{i=1}^{n_0}$ of n_0 space-filling design points, and running simulations to obtain their function values, $\mathbf{y}_{n_0} = \{f(x), x \in \mathbf{X}_{n_0}\}$. At each iteration, WIRBF first builds two global metamodels M_{S_n} with full information S_n and $M_{\widetilde{S}_n}$ with partial information $\widetilde{S}_n$. These two

---

Algorithm 1 An Adaptive Radial Basis Function Method using Weighted Improvement.

Inputs: Maximum allowed number of function evaluations $N_{\max}$; Initial sample size n_0.

Outputs: The best solution $x^\star_{N_{\max}}$ and its function value $f(x^\star_{N_{\max}})$ obtained by the algorithm.

Step 1. (Initialization and Space-Filling Designs)   Set $n = n_0$. Construct the space-filling design points $\mathbf{X}_{n_0} = \{x_i\}_{i=1}^{n_0}$, and set $\widetilde{\mathbf{X}}_{n_0} = \mathbf{X}_{n_0}$. Evaluate the function values of design points to define S_{n_0} and $\widetilde{S}_{n_0}$.

Step 2. (Build Global Metamodels)   Build two sets of RBF metamodels $M_{\widetilde{S}_n}$ and M_{S_n} with $\widetilde{S}_n$ and S_n, or update them using results in Section 3.1, and then employ model selection techniques in Section 3.2 to identify the most accurate model in each category.

Step 3. (Identify Candidate Points)

   3.1 (Global Search): Identify a candidate point x^g_{n+1} by solving $\max_{x \in \mathscr{X}} \mathrm{WI}(x \,|\, \widetilde{S}_n, f(x^\star_n), \eta)$.

   3.2 (Local Search): Initialize the local region $\mathscr{B}(\widetilde{x}^\star_n, \Delta_n)$, and determine a candidate point x^l_{n+1} by solving $\max_{x \in \mathscr{B}(\widetilde{x}^\star_n, \Delta_n)} \mathrm{WI}(x \,|\, S_n, f(\widetilde{x}^\star_n), \eta)$.

Step 4. (Simulation):   Solve $x^{new} = \arg\max_{x \in \{x^g_{n+1}, x^l_{n+1}\}} \mathrm{WI}(x \,|\, S_n, f(x^\star_n), \eta)$ and evaluate x^{new}.

Step 5. (Stopping Criterion):   Solve $\min_{x \in \mathbf{X}_n \cup x^{new}} f(x)$ to find $(x^\star_n, f(x^\star_n))$. If $n + |x^{new}| > N_{\max}$, then terminate the algorithm with the current best solution $x^\star_n$ as the final solution; otherwise initiate Step 6.

Step 6. (Unpromising iterations):   Declare the current iteration not promising if one of the following criteria is satisfied: (1) a small local region size $\Delta_n \le \varepsilon_\Delta$; (2) the number of function evaluations without introducing improvements exceeds N_f.

Step 7. (Restart):   If the current iteration is declared unpromising, then generate and evaluate a new set $\mathscr{C}^r$ of sample points and discard old samples by setting $\widetilde{\mathbf{X}}_{n+|\mathscr{C}^r|+1} = \mathscr{C}^r$; Otherwise, only set $\mathscr{C}^r = \emptyset$ and $\widetilde{\mathbf{X}}_{n+|\mathscr{C}^r|+1} = \widetilde{\mathbf{X}}_n \cup x^{new}$. Set $\mathbf{X}_{n+1+|\mathscr{C}^r|} = \mathbf{X}_n \cup x^{new} \cup \mathscr{C}^r$ and update $S_{n+|\mathscr{C}^r|+1}, \widetilde{S}_{n+|\mathscr{C}^r|+1}$, $n := n + |\mathscr{C}^r| + 1$ and go to Step 2.
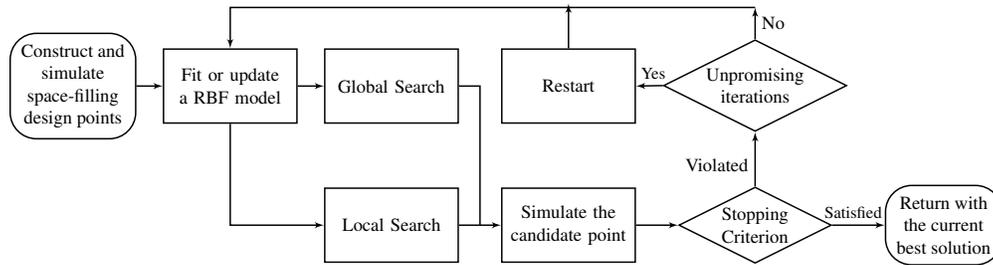

Figure 1: General structure of WIRBF.

metamodels serve different purposes. The model $M_{\widetilde{S}_n}$ is used to generate a candidate point x^g_{n+1} for global search. As $\widetilde{S}_n$ does not include sample points evaluated before last restart, Step 3.1 tends to generate sample points away from regions that were exhaustively explored. Meanwhile, since interpolating metamodels are considered in this study, M_{S_n} constructed with more information is generally more accurate than $M_{\widetilde{S}_n}$. To avoid being attracted by explored regions, M_{S_n} is used to generate a candidate point x^l_{n+1} within a local region $\mathscr{B}(\widetilde{x}^\star_n, \Delta_n) = \{x \in \mathscr{X} : \|x - \widetilde{x}^\star_n\|_\infty \le \Delta_n\}$. We use $\|\cdot\|_\infty$ for the convenience of generating local candidates within box-constrained $\mathscr{X} \cap \mathscr{B}(\widetilde{x}^\star_n, \Delta_n)$. In this study, WIRBF constructs an adaptive local region $\mathscr{B}(\widetilde{x}^\star_n, \Delta_n)$ with $n_{\mathscr{B}}$-nearest evaluated points $\{x^{(1)}, \ldots, x^{(n_{\mathscr{B}})}\} \subset \mathbf{X}_n$ to $\widetilde{x}^\star_n$, where $\Delta_n = \max_{i=1,\ldots,n_{\mathscr{B}}} \|x^{(i)} - \widetilde{x}^\star_n\|_\infty$.

One key element in WIRBF is the WI function that determines the function evaluational point to be the candidate point with large weighted improvement. In Step 3 of Algorithm 1, we apply the WI function

with different targets and RBF metamodels. In global search (Step 3.1), the WI function relies on $\widetilde{S}_n$ and $f(x_n^\star)$. Partial information $\widetilde{S}_n$ reduces the probability of being attracted by explored local optimal points. The global search focuses on improving the best solution found so far $x_n^\star$. Meanwhile, in the local search (Step 3.2), the WI function is based on S_n and $f(\widetilde{x}_n^\star)$. A more accurate model M_{S_n} based on S_n is helpful in finding promising points when the candidate point is restricted in the local region. In Step 4, we use all the information collected S_n and $f(x_n^\star)$ to decide the most promising point among these two candidate points.

Once the candidate point is evaluated in Step 4, WIRBF checks the stopping criterion. If the total number of function evaluations n exceeds a prefixed number $N_{\max}$, WIRBF terminates with the best solution found so far, $x_n^\star = \mathrm{argmin}_{x \in \mathbf{X}_n} f(x)$. Otherwise, WIRBF initiates Step 6 and spares computational efforts by interrupting unpromising iterations when the local region size $\Delta_n \leq \varepsilon_\Delta$ and no improvements are introduced within N_f number of function evaluations. After deciding to terminate unpromising iterations, WIRBF discards evaluated points in the partial information $\widetilde{S}_n$ and generate a set $\mathscr{C}^r$ of promising sample points by sequentially maximizing the Power function $\max_{x \in \mathscr{X}} \mathrm{P}^2_{\psi_n, \mathbf{X}_n}(x)$ for positive definite ψ or $\max_{x \in \mathscr{X}} \mathrm{P}^2_{\mathbf{C}_{\widehat{m}+n}, \mathbf{X}_n}(x)$ for conditionally positive definite ψ (see Section 3.3). This sampling scheme is equivalent to identifying points with maximum prediction error and is meant to explore the unknown regions.

3.1 Updating RBF Models

In WIRBF, RBF models are constantly updated with new evaluated points. Although guaranteed to be positive definite in theory, the interpolation matrix is prone to be ill-conditioned in practice because of clustered points around local optimal points. This is inefficient and error-prone to inverse the whole interpolation matrix each time. Mathematically, given an augmented RBF model M_{S_n}, let $S_{n+1} = S_n \cup \{(\hat{x}, f(\hat{x}))\}$, where $(\hat{x}, f(\hat{x}))$ is a new sampled point. Denote the weight vectors of M_{S_n} and $M_{S_{n+1}}$ by $\gamma_n \in \mathbb{R}^{\widehat{m}+n}$ and $\gamma_{n+1} \in \mathbb{R}^{\widehat{m}+n+1}$, respectively. With the Woodbury formula, we can compute γ_{n+1} without inverting the interpolation matrix of $M_{S_{n+1}}$,

$$\gamma_{n+1} = \mathbf{E}_{n+1}^{-1} \begin{pmatrix} \mathbf{C}_{\widehat{m}+n} & \widetilde{u}(\hat{x}) \\ \widetilde{u}(\hat{x})^\mathsf{T} & \psi(0) \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{\mathbf{y}}_n \\ f(\hat{x}) \end{pmatrix} = \mathbf{E}_{n+1}^{-1} \begin{pmatrix} \gamma_n + \mathsf{b}\mathbf{C}_{\widehat{m}+n}^{-1}\widetilde{u}(\hat{x})\widetilde{u}(\hat{x})^\mathsf{T}\gamma_n - \mathsf{b}\mathbf{C}_{\widehat{m}+n}^{-1}\widetilde{u}(\hat{x})f(\hat{x}) \\ -\mathsf{b}\widetilde{u}(\hat{x})^\mathsf{T}\gamma_n + \mathsf{b}f(\hat{x}) \end{pmatrix}, \quad (2)$$

where column vectors $u(\hat{x}) = \{\psi(\|\hat{x}-x_i\|)\}_{i=1}^n$, $v(\hat{x}) = \{\pi_j(\hat{x})\}_{j=1}^{\widehat{m}}$, $\widetilde{u}(\hat{x}) = (u(\hat{x})^\mathsf{T}, v(\hat{x})^\mathsf{T})^\mathsf{T}$, $\widetilde{\mathbf{y}}_n = \begin{pmatrix} \mathbf{y}_n \\ 0_{\widehat{m}} \end{pmatrix}$, $\gamma_n = \mathbf{C}_{\widehat{m}+n}^{-1}\widetilde{\mathbf{y}}_n$, and the scalar $\mathsf{b}^{-1} = \psi(0) - \widetilde{u}(\hat{x})^\mathsf{T}\mathbf{C}_{\widehat{m}+n}^{-1}\widetilde{u}(\hat{x})$. The permutation matrix $\mathbf{E}_{n+1}$ is formed by inserting the $(n+\widehat{m}+1)$th row of an identity matrix $\mathbf{I}_{(n+\widehat{m}+1)\times(n+\widehat{m}+1)}$ into its $(n+1)$th row. The computational complexity can be measured by the number of multiplications. The complexity of (2) is $O((n+\widehat{m})^2)$, while that of the algorithm of inverting the same interpolation matrix using Singular Value Decomposition (Golub and Van Loan 1996) is $O((n+\widehat{m})^3)$. The improvement is significant with a large n.

3.2 Model Selection

LOOCV has been applied as the model selection strategy in the literature. Brute force was used in Goel et al. (2007) to determine weights in the ensemble of surrogates. Rippa (1999) presented an analytical expression only for LOOCV on standard RBF models. In general, however, a closed form for LOOCV on augmented RBF models is not provided. By applying the rank-one update, we manage to derive a closed form for LOOCV on augmented RBF models, which accommodates a wider choice of radial basis functions compared to standard RBF models.

To apply LOOCV, the set of evaluated points $S_n = \{(x_i, y_i)\}_{i=1}^n$ is partitioned into n parts. For the ith part, we fit a model $M_{S_{-i}}$ with evaluated points $S_{-i} = S_n \setminus \{(x_i, y_i)\}$ and measure its accuracy with the ith evaluated point. We can compute the LOOCV error $\mathrm{CV}_n(M_{S_n})$ of the augmented RBF model M_{S_n} as follows

$$\mathrm{CV}_n(M_{S_n}) = \frac{1}{n}\sum_{i=1}^n \left|y_i - M_{S_{-i}}(x_i)\right| = n^{-1}\sum_{i=1}^n \left|\gamma_i[\mathbf{C}_{\widehat{m}+n}^{-1}]_{i,i}^{-1}\right|.$$

where $[\mathsf{M}]_{i,i}$ is the (i,i)th element of matrix M, $\gamma \in \mathbb{R}^{\widehat{\mathsf{m}}+n}$ is the weight vector of M_{S_n}, and $\mathbf{C}_{\widehat{\mathsf{m}}+n} = \begin{pmatrix} \Psi_n & \Pi \\ \Pi^\mathsf{T} & 0_{\widehat{\mathsf{m}} \times \widehat{\mathsf{m}}} \end{pmatrix}$ is the interpolation matrix of M_{S_n}.

3.3 Error Estimates of RBF Models

In this section, we propose a pointwise error estimate for the prediction of a standard RBF model using leave-one-out cross validation. Suppose that a standard RBF model M_{S_n} interpolates an unknown function f at a set S_n of evaluated points, our goal is to estimate $|M_{S_n}(\hat{x}) - f(\hat{x})|$ for all $\hat{x} \in \mathscr{X}$. To proceed with this goal, we need consider a certain class of f, which is formalized in Assumption 1.

Assumption 1 The unknown function f is assumed to be smooth and belong to a Reproducing Kernel Hilbert Space (RKHS) $\mathscr{H}$ generated by a reproducing kernel K, where K is also the radial basis function used in M_{S_n}.

Excellent references on RKHS include Aronszajn (1950) and Wendland (2005). Assumption 1 allows us to mathematically characterize the smoothness of any function $g \in \mathscr{H}$ by $\|g\|_{\mathscr{H}}$. For any $g \in \mathscr{H}$ and $\hat{x} \in \mathscr{X}$, by applying Cauchy-Schwarz inequality,

$$|g(\hat{x})| = \left|\langle g, K(\cdot,\hat{x})\rangle_{\mathscr{H}}\right| \leq \|g\|_{\mathscr{H}} \|K(\cdot,\hat{x})\|_{\mathscr{H}} = \|g\|_{\mathscr{H}}\sqrt{\langle K(\cdot,\hat{x}), K(\cdot,\hat{x})\rangle_{\mathscr{H}}} = \|g\|_{\mathscr{H}}\sqrt{K(\hat{x},\hat{x})}.$$

The upper bound of $|g(\hat{x})|$ is proportional to $\|g\|_{\mathscr{H}}$. In particular, when radial kernels are used, $K(\hat{x},\hat{x})$ is a constant for all $\hat{x} \in \mathscr{X}$. Thus $\|\cdot\|_{\mathscr{H}}$ is a plausible measure of smoothness of a function. Moreover, $\|g\|_{\mathscr{H}}$ can be computed efficiently for any $g \in \mathscr{H}$. Wendland (2005) gives the RKHS norm of a standard RBF model M_{S_n} as $\|M_{S_n}\|_{\mathscr{H}}^2 = \alpha^\mathsf{T}\Psi_n\alpha$.

To estimate the prediction error of $M_{S_n}(\hat{x})$, we treat its true function value $f(\hat{x})$ as a variable $\hat{y} = f(\hat{x})$ and construct a new metamodel $M_{\widehat{S}_{n+1}}$ with $\widehat{S}_{n+1} = S_n \cup \{(\hat{x},\hat{y})\}$. Assuming a smooth f, we may expect a reasonable metamodel to be smooth as well and thus focus on its smoothness measure $\|M_{\widehat{S}_{n+1}}\|_{\mathscr{H}}^2$. The model $M_{\widehat{S}_{n+1}}$ is defined by $M_{\widehat{S}_{n+1}}(x) = \sum_{i=1}^{n}\alpha_i'\psi(\|x-x_i\|) + \alpha_{n+1}'\psi(\|x-\hat{x}\|)$. The unknown weights $\{\alpha_i'\}_{i=1}^{n+1}$ are determined by the interpolation conditions $\begin{pmatrix} \Psi_n & u(\hat{x}) \\ u(\hat{x})^\mathsf{T} & \psi(0) \end{pmatrix}\alpha' = \begin{pmatrix} \mathbf{y}_n \\ \hat{y} \end{pmatrix}$, where $\Psi_n = \{\psi(\|x_i - x_j\|)\}_{i,j=1}^{n}$ and $u(\hat{x}) = \{\psi(\|\hat{x} - x_i\|)\}_{i=1}^{n}$. Since $M_{\widehat{S}_{n+1}} \in \mathscr{H}$, its norm is given by

$$\begin{aligned}
\|M_{\widehat{S}_{n+1}}\|_{\mathscr{H}}^2 &= (\alpha')^\mathsf{T}\begin{pmatrix} \Psi_n & u(\hat{x}) \\ u(\hat{x})^\mathsf{T} & \psi(0) \end{pmatrix}\alpha' = \begin{pmatrix} \mathbf{y}_n & \hat{y} \end{pmatrix}\begin{pmatrix} \Psi_n & u(\hat{x}) \\ u(\hat{x})^\mathsf{T} & \psi(0) \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{y}_n \\ \hat{y} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{y}_n & \hat{y} \end{pmatrix}\begin{pmatrix} \Psi_n^{-1} + \mathsf{b}\Psi_n^{-1}u(\hat{x})u(\hat{x})^\mathsf{T}\Psi_n^{-1} & -\mathsf{b}\Psi_n^{-1}u(\hat{x}) \\ -\mathsf{b}u(\hat{x})^\mathsf{T}\Psi_n^{-1} & \mathsf{b} \end{pmatrix}\begin{pmatrix} \mathbf{y}_n \\ \hat{y} \end{pmatrix} \\
&= \|M_{S_n}\|_{\mathscr{H}}^2 + (\psi(0) - u(\hat{x})^\mathsf{T}\Psi_n^{-1}u(\hat{x}))^{-1}(M_{S_n}(\hat{x}) - \hat{y})^2.
\end{aligned} \tag{3}$$

By specifying a target value $\hat{y}$ as an approximation of the optimal objective value, Gutmann (2001) selects a candidate point $\hat{x}$ that minimizes (3), which implies that the true function value $f(\hat{x})$ is most like to be $\hat{y}$ since the resulted metamodel $M_{\widehat{S}_{n+1}}$ is the smoothest. In the following, we interpret (3) from another standpoint and derive pointwise error estimates for standard RBF models. The basic idea is to estimate the upper bound of $\|M_{\widehat{S}_{n+1}}\|_{\mathscr{H}}^2$ that in turn limits the possible value of $\hat{y}$. First, (3) implies that for any $\hat{x} \notin \mathbf{X}_n$, $(\psi(0) - u(\hat{x})^\mathsf{T}\Psi_n^{-1}u(\hat{x}))^{-1} > 0$ since it is one diagonal element of the inverse of a positive definite interpolation matrix. Thus, we obtain the following proposition,

Proposition 1 Suppose $\{S_n\}_{n\in\mathbb{N}}$ is an increasing sequence of distinct evaluated points in $\mathscr{X}$, that is, $S_n \subseteq S_{n+1}$ for any $n \in \mathbb{N}$. Then $\|M_{S_n}\|_{\mathscr{H}} \leq \|M_{S_{n+1}}\|_{\mathscr{H}} \leq \cdots \leq \|f\|_{\mathscr{H}}$.

Clearly, $\|M_{\widehat{S}_{n+1}}\|_{\mathscr{H}} \in \mathscr{H}$ should also be upper bounded by $\|f\|_{\mathscr{H}}$,

$$\|M_{\widehat{S}_{n+1}}\|_{\mathscr{H}}^2 = \|M_{S_n}\|_{\mathscr{H}}^2 + (\psi(0) - u(\hat{x})^\mathsf{T}\Psi_n^{-1}u(\hat{x}))^{-1}(M_{S_n}(\hat{x}) - \hat{y})^2 \leq \|f\|_{\mathscr{H}}^2.$$

Since we set $\hat{y} = f(\hat{x})$, we obtain an error estimate of $M_{S_n}(\hat{x})$ by solving the inequality for $\hat{y}$,

$$|f(\hat{x}) - M_{S_n}(\hat{x})|^2 \leq (\psi(0) - u(\hat{x})^\mathsf{T}\Psi_n^{-1}u(\hat{x}))(\|f\|_{\mathscr{H}}^2 - \|M_{S_n}\|_{\mathscr{H}}^2). \tag{4}$$

This error bound has several implications: First, Corollary 18.1 in Fasshauer (2007) suggests that $\|f\|_{\mathscr{H}}^2 - \|M_{S_n}\|_{\mathscr{H}}^2 = \|f - M_{S_n}\|_{\mathscr{H}}^2$. Meanwhile, Theorem 10.3 in Wendland (2005) states that convergence in $\|\cdot\|_{\mathscr{H}}$ implies pointwise convergence, that is, $\lim_{n\to\infty} M_{S_n}(\hat{x}) = f(\hat{x})$ for all $\hat{x} \in \mathscr{X}$. Thus $\|f - M_{S_n}\|_{\mathscr{H}}$ can be viewed as a measure of the overall model accuracy of M_{S_n}, while $\psi(0) - u(\hat{x})^\mathsf{T}\Psi_n^{-1}u(\hat{x})$ acts as a distributor of the residual over the entire region $\mathscr{X}$. We therefore hypothesize that $\|f - M_{S_n}\|_{\mathscr{H}}$ may be estimated with model validation techniques, such as cross validation.

Next, the pointwise error estimate may be conservative if the upper bound $\|f\|_{\mathscr{H}}^2$ of $\|M_{S_n}\|_{\mathscr{H}}^2$ is used in (4). A conservative error estimate will promote metamodel-based optimization algorithms to search globally. Therefore, we propose to replace $\|f\|_{\mathscr{H}}^2$ with an estimated upper bound of $\|M_{\hat{S}_{n+1}}\|_{\mathscr{H}}^2$ so that the estimated error bound is probabilistically satisfied. Finally, (3) characterizes the exact relationship between $\|M_{\hat{S}_{n+1}}\|_{\mathscr{H}}^2$ and $\|M_{S_n}\|_{\mathscr{H}}^2$. If k-fold ($k < n$) cross validation is used, it is rather difficult to derive an exact relationship between $\|M_{\hat{S}_{n+1-k}}\|_{\mathscr{H}}^2$ and $\|M_{S_n}\|_{\mathscr{H}}^2$ similar to (3).

Based on the above justifications, we propose to estimate $\|M_{\hat{S}_{n+1}}\|_{\mathscr{H}}^2$ via leave-one-out cross validation. $\|M_{\hat{S}_{n+1}}\|_{\mathscr{H}}^2 - \|M_{S_n}\|_{\mathscr{H}}^2$ can be seen as the contribution of an point $\hat{x}$ to the overall accuracy of a metamodel M_{S_n}. For standard RBF models, $\|M_{\hat{S}_{n+1}}\|_{\mathscr{H}}^2 - \|M_{S_n}\|_{\mathscr{H}}^2 \geq 0$ for all $\hat{x} \in \mathscr{X}$. In order to estimate $\|M_{\hat{S}_{n+1}}\|_{\mathscr{H}}^2$, we make the following assumption relating $\|M_{\hat{S}_{n+1}}\|_{\mathscr{H}}^2 - \|M_{S_n}\|_{\mathscr{H}}^2$ with $\{\|M_{S_n}\|_{\mathscr{H}}^2 - \|M_{S_{-i}}\|_{\mathscr{H}}^2\}_{i=1}^n$.

Assumption 2 We treat $\|M_{\hat{S}_{n+1}}\|_{\mathscr{H}}^2 - \|M_{S_n}\|_{\mathscr{H}}^2$ and $\{\|M_{S_n}\|_{\mathscr{H}}^2 - \|M_{S_{-i}}\|_{\mathscr{H}}^2\}_{i=1}^n$ as if they were independent random variables from the same distribution.

Proposition 2 The estimated error bound of a standard RBF model M_{S_n} is given by

$$\mathbb{P}\left[|f(\hat{x}) - M_{S_n}(\hat{x})|^2 \leq \mathsf{P}_{\Psi_n,\mathbf{X}_n}^2(\hat{x})\rho\left(\alpha, \Psi_n^{-1} \mid \eta\right)\right] = \eta, \tag{5}$$

where $\rho\left(\alpha, \Psi_n^{-1} \mid \eta\right)$ is estimated from $\{\|M_{S_n}\|_{\mathscr{H}}^2 - \|M_{S_{-i}}\|_{\mathscr{H}}^2\}_{i=1}^n$ and $\mathsf{P}_{\Psi_n,\mathbf{X}_n}^2(\hat{x}) = \psi(0) - u(\hat{x})^\mathsf{T}\Psi_n^{-1}u(\hat{x})$.

In this study, we simply use the inverse of empirical cumulative distribution function, that is, $\rho\left(\alpha, \Psi_n^{-1} \mid \eta\right)$ is the η-quantile $Q_\eta\left[\{\|M_{S_n}\|_{\mathscr{H}}^2 - \|M_{S_{-i}}\|_{\mathscr{H}}^2\}_{i=1}^n\right]$ with $\eta \in [0,1]$. By applying results from Section 3.2, we obtain

$$\|M_{S_n}\|_{\mathscr{H}}^2 - \|M_{S_{-i}}\|_{\mathscr{H}}^2 = \alpha_i^2[\Psi_n^{-1}]_{i,i}^{-1}.$$

It is worth noting that $\mathsf{P}_{\Psi_n,\mathbf{X}_n}^2(\hat{x}) = 0$ if and only if $\hat{x} \in \mathbf{X}_n$ and $\mathsf{P}_{\Psi_n,\mathbf{X}_n}^2(\hat{x}) > 0$ for all $\hat{x} \notin \mathbf{X}_n$. Therefore, (5) still preserves the property that the prediction error at evaluated points is zero. The complexity of (5) is $O(n^2)$. Most computational efforts are spent in computing $u(\hat{x})^\mathsf{T}\Psi_n^{-1}u(\hat{x})$. In the case of local radial basis functions such as GS, the complexity can be further reduced by exploiting the sparse structure of $u(\hat{x})$ and Ψ_n^{-1}. For simplicity, we only present the estimated error bound of augmented RBF models:

$$\mathbb{P}\left[|f(\hat{x}) - M_{S_n}(\hat{x})|^2 \leq \mathsf{P}_{\mathbf{C}_{\hat{\mathsf{m}}+n},\mathbf{X}_n}^2(\hat{x})\rho\left(\gamma, \mathbf{C}_{\hat{\mathsf{m}}+n}^{-1} \mid \eta\right)\right] = \eta,$$

where $\mathsf{P}_{\mathbf{C}_{\hat{\mathsf{m}}+n},\mathbf{X}_n}^2(\hat{x}) = \psi(0) - \widetilde{u}(\hat{x})^\mathsf{T}\mathbf{C}_{\hat{\mathsf{m}}+n}^{-1}\widetilde{u}(\hat{x})$ and $\rho\left(\gamma, \mathbf{C}_{\hat{\mathsf{m}}+n}^{-1} \mid \eta\right) = Q_\eta\left[\left\{\gamma_i^2[\mathbf{C}_{\hat{\mathsf{m}}+n}^{-1}]_{i,i}^{-1}\right\}_{i=1}^n\right]$. The complexity of this estimated error bound is $O((\hat{\mathsf{m}}+n)^2)$.

To demonstrate the effectiveness of the error estimate, in Figure 2, we compare it with the 95% confidence interval for a kriging model on a randomly-generated function.
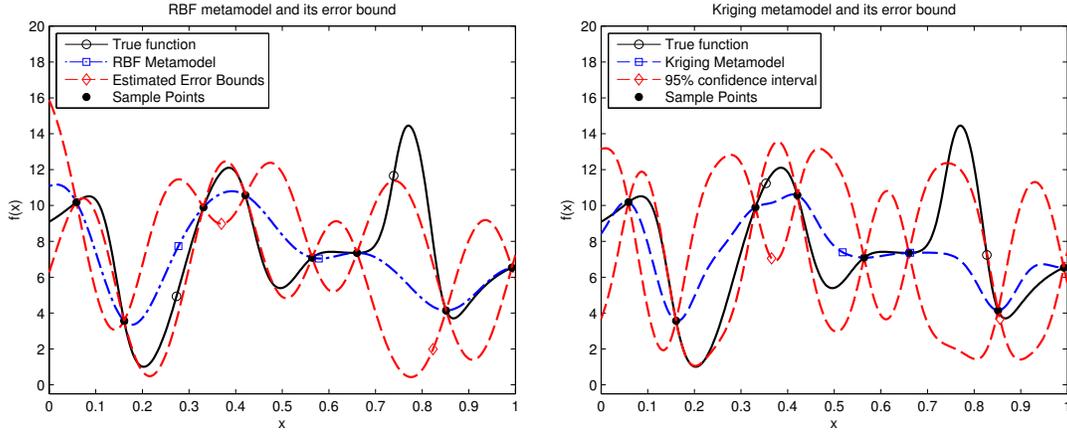
Figure 2: The left figure shows the estimated error bounds of RBF models with $\eta = 0.95$ and $\psi(r) = \exp(-r^2/\theta)$, where θ is chosen by minimizing the LOOCV error $\mathrm{CV}_n(M_{S_n})$; the right figure shows 95% confidence interval of kriging models with Gaussian correlation functions.

3.4 Weighted Improvement

Given a set of sample points $S_n = \{\mathbf{X}_n, \mathbf{y}_n\}$ and a target value τ^*, the weighted improvement $\mathsf{WI}(\hat{x} \mid S_n, \tau^*)$ at an unknown point $\hat{x}$ primarily relies on two elements: (a) the improvement $I_{\hat{x}} = \max(\tau^* - \hat{y}, 0) = (\tau^* - \hat{y})^+$, where the variable $\hat{y}$ represents the unknown $f(\hat{x})$; (b) a weight function $W(\hat{y})$ for each possible $I_{\hat{x}} = (\tau^* - \hat{y})^+$. We consider the following weight function,

$$W(\hat{y}) = \frac{\|M_{S_n}\|_{\mathscr{H}}}{\pi \mathsf{P}_{\Psi_n, \mathbf{X}_n}(\hat{x})} \times \|M_{\widehat{S}_{n+1}}\|_{\mathscr{H}}^{-2} = \frac{\pi^{-1} \|M_{S_n}\|_{\mathscr{H}} \mathsf{P}_{\Psi_n, \mathbf{X}_n}(\hat{x})}{\|M_{S_n}\|_{\mathscr{H}}^2 \mathsf{P}_{\Psi_n, \mathbf{X}_n}^2(\hat{x}) + (\hat{y} - M_{S_n}(\hat{x}))^2},$$

where $\|M_{S_n}\|_{\mathscr{H}}, M_{S_n}(\hat{x})$, and $\mathsf{P}_{\Psi_n, \mathbf{X}_n}(\hat{x})$ are all constants given $\hat{x}$, ψ and S_n; Our choice of $W(\hat{y})$ exhibits three properties: (a) It monotonically decreases as $\hat{y}$ deviates from $M_{S_n}(\hat{x})$, implying the metamodel $M_{\widehat{S}_{n+1}}$ becomes bumpier; (b) $W(\hat{y}) \geq 0$, for any $\hat{y} \in \mathbb{R}$; (c) $\int_{-\infty}^{+\infty} W(\hat{y}) \, \mathrm{d}\hat{y} = 1$. A closer look at $W(\hat{y})$ reveals that it can be treated as the probability density function of a Cauchy random variable $\hat{y}$ with two robust measures: median $M_{S_n}(\hat{x})$ and median absolute deviation $\|M_{S_n}\|_{\mathscr{H}} \mathsf{P}_{\Psi_n, \mathbf{X}_n}(\hat{x})$.

With the two key elements defined, we now compute $\mathsf{WI}(\hat{x} \mid S_n, \tau^*)$ by integrating the weighted improvement $I_{\hat{x}}$ over all possible $\hat{y}$,

$$\mathsf{WI}(\hat{x} \mid S_n, \tau^*) = \left(\int W(\hat{y}) \, \mathrm{d}\hat{y} \right)^{-1} \int W(\hat{y}) I_{\hat{x}} \, \mathrm{d}\hat{y}.$$

However, the integration would be undefined for unconstrained $\hat{y}$. Therefore, in our weighted improvement procedure, we restrict the possible $\hat{y}$ within the estimated error bound (5), that is, $\hat{y} \in [\tilde{f}_l(\hat{x}), \tilde{f}_u(\hat{x})]$, where $\tilde{f}_l(\hat{x}) = M_{S_n}(\hat{x}) - \sqrt{\mathsf{P}_{\Psi_n, \mathbf{X}_n}^2(\hat{x}) \rho\left(\alpha, \Psi_n^{-1} \mid \eta\right)}$, and $\tilde{f}_u(\hat{x}) = M_{S_n}(\hat{x}) + \sqrt{\mathsf{P}_{\Psi_n, \mathbf{X}_n}^2(\hat{x}) \rho\left(\alpha, \Psi_n^{-1} \mid \eta\right)}$. Such a restriction brings simplification, yielding a closed form for the WI function. Meanwhile, this restriction can be adjusted by changing the size of $[\tilde{f}_l(\hat{x}), \tilde{f}_u(\hat{x})]$ with η. We obtain the WI at $\hat{x}$,

$$\mathsf{WI}(\hat{x} \mid S_n, \tau^*, \eta) = \|M_{S_n}\|_{\mathscr{H}} \mathsf{P}_{\Psi_n, \mathbf{X}_n}(\hat{x}) \left(\frac{\ln(1 + \mathsf{v}^2) - \ln(1 + \mathsf{u}^2)}{2(\tan^{-1}(\mathsf{u}) - \tan^{-1}(\mathsf{v}))} + \frac{\tau^* - M_{S_n}(\hat{x})}{\|M_{S_n}\|_{\mathscr{H}} \mathsf{P}_{\Psi_n, \mathbf{X}_n}(\hat{x})} \right), \tag{6}$$

where $u = \min\left(\frac{\tau^* - M_{S_n}(\hat{x})}{\|M_{S_n}\|_{\mathscr{H}}P_{\Psi_n,\mathbf{X}_n}(\hat{x})}, \frac{\sqrt{\rho(\alpha,\Psi_n^{-1}|\eta)}}{\|M_{S_n}\|_{\mathscr{H}}}\right)$ and $v = \min\left(\frac{\tau^* - M_{S_n}(\hat{x})}{\|M_{S_n}\|_{\mathscr{H}}P_{\Psi_n,\mathbf{X}_n}(\hat{x})}, -\frac{\sqrt{\rho(\alpha,\Psi_n^{-1}|\eta)}}{\|M_{S_n}\|_{\mathscr{H}}}\right)$. The WI function for augmented RBF models could be obtained by simply replacing $P^2_{\Psi_n,\mathbf{X}_n}(\hat{x})$ and $\rho\left(\alpha,\Psi_n^{-1}|\eta\right)$ with their counterparts, $P^2_{\mathbf{C}_{\hat{\mathsf{m}}+n},\mathbf{X}_n}(\hat{x})$ and $\rho\left(\gamma,\mathbf{C}_{\hat{\mathsf{m}}+n}^{-1}|\eta\right)$.

To better comprehend the WI function, we plot (6) for a randomly-generated function and compare it with the EI function in Figure 3. The two metamodels are constructed with the same set of sample points. For the RBF model, we choose a cubic $\psi(r) = r^3$ and $\eta = 0.25, 0.95$. The kriging model uses a Gaussian correlation function. The target value τ^* is highlighted with the horizontal dashed line. As
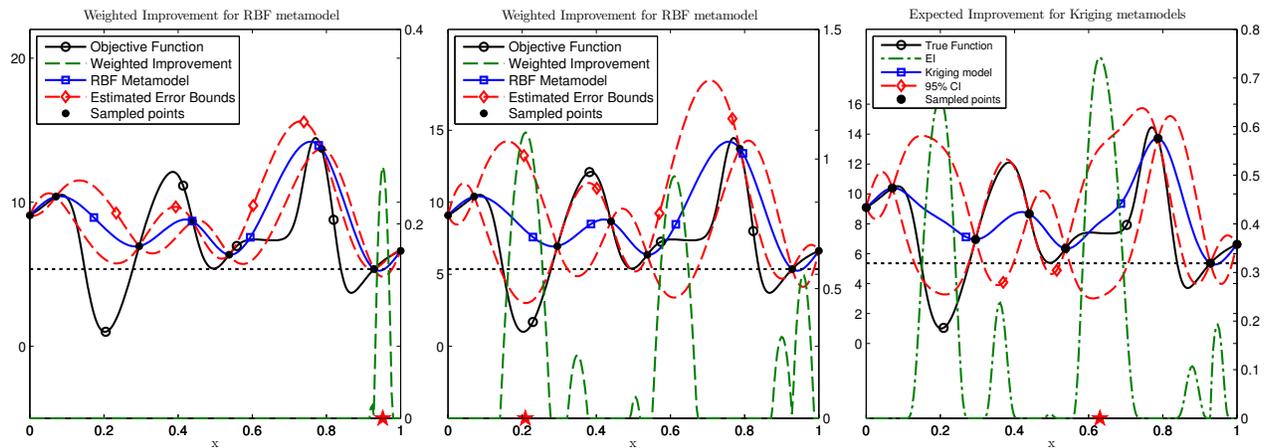

Figure 3: The left figure shows the WI function with $\eta = 0.25$; the middle figure uses $\eta = 0.95$; the right figure illustrates the EI function of a kriging model. The star denotes the point with maximum WI or EI.

depicted in the left and middle panel of Figure 3, a lower η promotes local search as the estimated error bound $[\tilde{f}_l(\hat{x}), \tilde{f}_u(\hat{x})]$, $\hat{x} \in \mathscr{X}$ of the RBF model is narrower. When $\eta = 0.95$, the WI function identifies similar promising regions (six peaks of the WI function) as the EI function in kriging does.

4 NUMERICAL EXPERIMENTS

In this section, we compare WIRBF with four alternative approaches: EGO (Jones et al. 1998), rbfSolve (Gutmann 2001), MADS-DACE (Audet and Dennis 2006), and MLMSRBF (Regis and Shoemaker 2007b). We use EGO and rbfSolve implemented in the TOMLAB optimization environment, MADS-DACE implemented in the NOMADm software package, and a Matlab version of MLMSRBF implemented by its author.

4.1 Experimental Setup

We consider a set $\mathscr{T}$ of 45 different test problems whose dimensions range from 2 to 15 with the median dimension being 8. Most problems have been used in the literature Gutmann (2001) and Regis and Shoemaker (2007b) to compare different global optimization algorithms. Seven of $\mathscr{T}$ are selected from (Dixon and Szegö 1978). Besides, we include Beale ($d = 2$), Zakharov$_d$ ($d = 5, 10$) and Powell's singular function Powell$_d$ ($d = 8, 12$) from the literature. Another subset of $\mathscr{T}$ comprises deterministic test problems that attempt to mimic stochastic simulations with high-frequency and low-frequency oscillations in the form of sine and cosine functions, including Ackley$_d$ ($d = 2, 5, 10, 15$), Schwefel$_d$ ($d = 2$), Griewank$_d$ ($d = 15$). Aside from these standard test problems, we also include 26 random instances of the Schoen (1993) test

problems, labeled as Schoen$_{d,20}$, Schoen$_{d,100}$, $d = 3, \ldots, 15$. Schoen$_{d,20}$ has more than 10 local optimal points, and Schoen$_{d,100}$ exhibits more than 50 local optimal points, four of which are steep local minima.

To examine the robustness, we test all algorithms on each problem with 50 replications using $N_{max} = 850$ function evaluations. In each replication, each algorithm starts with $n_0 = 2(d+1)$ initial points randomly generated by Latin Hypercube Sampling implemented in the Matlab command `lhsdesign()`. We leave all parameters in MADS-DACE, EGO, MLMSRBF and rbfSolve at default values in this study.

The implementation of WIRBF scales the feasible region of each problem into $[0,1]^d$, which is also adopted by the other four algorithms. The WIRBF implementation demonstrated here relies on parameters as follows: $\eta = 0.6$, $|\mathscr{C}^r| = 2(d+1)$, $n_{\mathscr{B}} = 2d$, $N_f = \max(8, d)$, $\varepsilon_\Delta = 10^{-4}\sqrt{d}$. Our computational experience shows that WIRBF also works well when it approximates the optimal point of the WI function with the best one selected from a limited number of points. Specifically, we randomly generate $N_g = 500d$ points in Step 3.1 and $N_l = 300d$ in Step 3.2. Furthermore, we consider three variants of WIRBF: In WIRBF-AUTO, all five basis functions in Table 1 are used in the model selection, where the shape parameter in MQ, IMQ and GS is chosen by minimizing the LOOCV error; WIRBF-CU and WIRBF-TPS respectively uses CU and TPS to construct RBF models throughout the iterations.

4.2 Experimental Results

We now present computational results using data profile (Moré and Wild 2009). This profiling technique depends on three components: a set $\mathscr{A}$ of optimization algorithms, a set $\mathscr{T}$ of test problems, and a convergence test. The derivative-free convergence test is given by

$$f(x_0) - f(x) \geq (1 - \varepsilon_{ct})(f(x_0) - f_L),$$

where $\varepsilon_{ct} > 0$ is a tolerance, x_0 is the starting point for the problem and is set to be the best solution found after evaluating n_0 space-filling design points, and f_L is the smallest function value obtained by any tested algorithm within N_{max} function evaluations and 50 runs.

Figure 4 shows the data profiles for different accurate levels of $\varepsilon_{ct} = \{10^{-1}, 10^{-5}\}$. WIRBF-AUTO generally performs best. The performance of the variants of WIRBF are similar when an low accurate level $\varepsilon_{ct} = 10^{-1}$ is demanded, while the differences among their performances increases at the accuracy level $\varepsilon_{ct} = 10^{-5}$. Given the equivalent 30 simplex gradient ($30(d+1)$ function evaluations), WIRBF-AUTO, WIRBF-CU and WIRBF-TPS respectively solve 82%, 80%, and 82% of problems to $\varepsilon_{ct} = 10^{-1}$, and 41%, 35%, and 31% of problems to $\varepsilon_{ct} = 10^{-5}$, while MLMSRBF, rbfSolve, EGO, and MADS-DACE respectively solve 72%, 54%, 34%, and 52% of problems to $\varepsilon_{ct} = 10^{-1}$, and 21%, 8%, 12% and 7% of problems to $\varepsilon_{ct} = 10^{-5}$.

We are also interested in the type of radial basis function selected by WIRBF-AUTO. Figure 5a suggests that the LOOCV-based model selection strategy favors MQ for small sample size, while TPS becomes the top choice for large sample size. It is surprising to observe a low percentage of choosing CU despite the good performance of WIRBF-CU. Figure 5b presents the mean normalized LOOCV error of different RBF models for different number of sample points across the experiments. A higher value indicates a lower LOOCV error. We can see that the performance of IMQ, TPS and CU increases with the sample size relative to other basis functions, but CU is seldom the best thus is not favored according to the criterion of minimizing LOOCV error.

5 CONCLUSIONS

In this paper, we have presented an adaptive radial basis function method using weighted improvement. We first propose to construct RBF models using rank-one update and then derive a closed form of LOOCV errors of RBF models. We also propose a new error estimate for RBF interpolation models via leave-one-out cross validation. Previous works mainly use the minimum distance to evaluated points as an indicator of the prediction uncertainty. Computational results show that the three variants of WIRBF outperforms
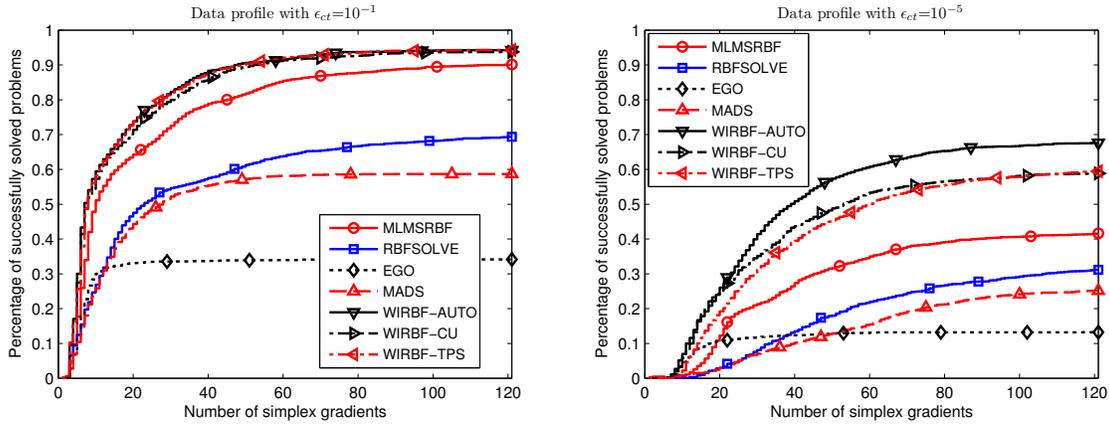
Figure 4: Data profiles for algorithms $\mathscr{A}$ on test problems $\mathscr{T}$.
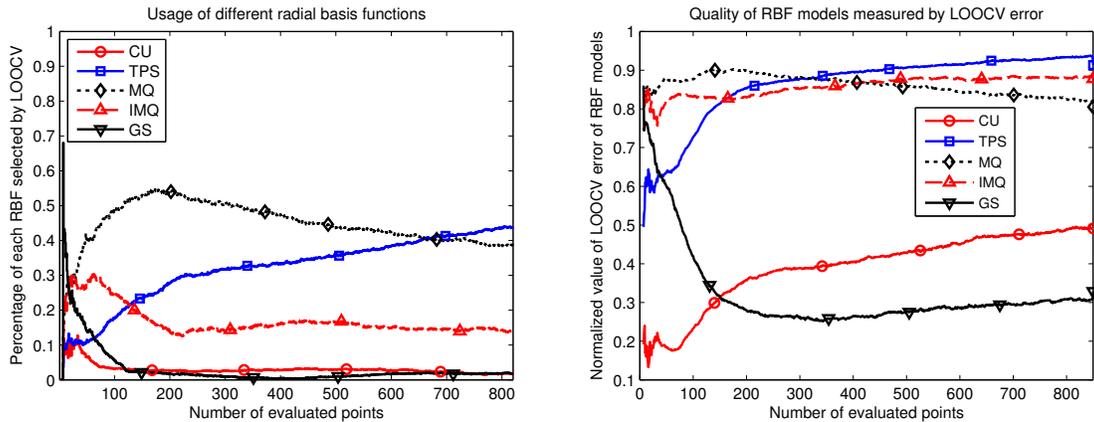

Figure 5: Radial basis functions selected by LOOCV with 50 runs and 45 problems.

existing algorithms. We also find out that LOOCV-based model selection strategy favors MQ for small sample size, while TPS becomes the top choice for a large sample size, and the performance of IMQ, CU and TPS generally increases with the sample size relative to other basis functions.

ACKNOWLEDGMENTS

REFERENCES

Aronszajn, N. 1950. "Theory of reproducing kernels". *Trans. Amer. Math. Soc* 68 (3): 337–404.

Audet, C., and J. E. Dennis, Jr.. 2006. "Mesh Adaptive Direct Search Algorithms for Constrained Optimization". *SIAM Journal on Optimization* 17:188–217.

Björkman, M., and K. Holmström. 2000. "Global Optimization of Costly Nonconvex Functions using Radial Basis Functions". *Optimization and Engineering* 1 (4): 373–397.

Booker, A., J. Dennis, P. Frank, D. Serafini, V. Torczon, and M. Trosset. 1999. "A Rigorous Framework for Optimization of Expensive Functions by Surrogates". *Structural and Multidisciplinary Optimization* 17 (1): 1–13.

Dixon, L., and G. Szegö. 1978. "The Global Optimization Problem: an Introduction". *Towards Global Optimization* 2:1–15.

Fasshauer, G. 2007. *Meshfree Approximation Methods with Matlab*. Interdisciplinary Mathematical Sciences Series. World Scientific.

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

Golub, G., and C. Van Loan. 1996. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press.

Gutmann, H.-M. 2001. "A Radial Basis Function Method for Global Optimization". *Journal of Global Optimization* 19:201–227.

Hardy, R. 1971. "Multiquadric Equations of Topography and Other Irregular Surfaces". *Journal of Geophysical Research* 76 (8): 1905–1915.

Jakobsson, S., M. Patriksson, J. Rudholm, and A. Wojciechowski. 2010. "A Method for Simulation based Optimization using Radial Basis Functions". *Optimization and Engineering* 11 (4): 501–532.

Jones, D. 2001. "A Taxonomy of Global Optimization Methods based on Response Surfaces". *Journal of Global Optimization* 21 (4): 345–383.

Jones, D., M. Schonlau, and W. Welch. 1998. "Efficient Global Optimization of Expensive Black-box Functions". *Journal of Global Optimization* 13 (4): 455–492.

Moré, J., and S. Wild. 2009. "Benchmarking Derivative-free Optimization Algorithms". *SIAM Journal on Optimization* 20 (1): 172–191.

Powell, M. 1992. "The theory of Radial Basis Function Approximation". *Advances in Numerical Analysis* 2:105–210.

Regis, R. 2011. "Stochastic Radial Basis Function Algorithms for Large-scale Optimization Involving Expensive Black-box Objective and Constraint Functions". *Computers & Operations Research* 38 (5): 837–853.

Regis, R., and C. Shoemaker. 2007a. "Improved Strategies for Radial Basis Function Methods for Global Optimization". *Journal of Global Optimization* 37 (1): 113–135.

Regis, R., and C. Shoemaker. 2007b. "A Stochastic Radial Basis Function Method for the Global Optimization of Expensive Functions". *INFORMS Journal on Computing* 19:457–509.

Regis, R. G., and C. A. Shoemaker. 2005. "Constrained Global Optimization of Expensive Black Box Functions Using Radial Basis Functions". *Journal of Global Optimization* 31 (1): 153–171.

Rippa, S. 1999. "An Algorithm for Selecting a Good Value for the Parameter c in Radial Basis Function Interpolation". *Advances in Computational Mathematics* 11 (2): 193–210.

Schoen, F. 1993. "A Wide Class of Test Functions for Global Optimization". *Journal of Global Optimization* 3:133–137.

Wendland, H. 2005. *Scattered Data Approximation*, Volume 2. Cambridge University Press.

AUTHOR BIOGRAPHIES

YIBO JI is a Ph.D. candidate in the Department of Industrial and Systems Engineering, National University of Singapore. He received his B.Eng. degree in Industrial Engineering at the School of Management and Engineering of Nanjing University in 2009. His research interests include simulation optimization, and metamodel-based global optimization problems. His email address is jiyibo@nus.edu.sg.

SUJIN KIM is an assistant professor in the Department of Industrial and Systems Engineering at the National University of Singapore. She received her Ph.D. degree in Operations Research from Cornell University in 2006. Before she joined the National University of Singapore, she was a visiting assistant professor in the Department of Industrial Engineering at Purdue University. Her research concerns simulation methodology and stochastic simulation-based optimization, with applications in electric power and health service systems. Her e-mail address is iseks@nus.edu.sg.