

## **Application of Support Vector Machines to Accelerate the Solution Speed of Metaheuristic Algorithms**

### Author

Yang, Shiyou, H. Liu, Qing, Lu, Junwei, o, S., Ni, Guangzheng, Ni, Peihong, Xiong, Suming

### Published

2009

### Journal Title

IEEE Transactions on Magnetics

### DOI

[10.1109/TMAG.2009.2012690](https://doi.org/10.1109/TMAG.2009.2012690)

### Rights statement

© 2009 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

### Downloaded from

<http://hdl.handle.net/10072/29648>

### Link to published version

<http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=20>

### Griffith Research Online

<https://research-repository.griffith.edu.au>

# Application of Support Vector Machines to Accelerate the Solution Speed of Metaheuristic Algorithms

Shiyou Yang<sup>1</sup>, Qing H. Liu<sup>2</sup>, Junwei Lu<sup>3</sup>, S. L. Ho<sup>4</sup>, Guangzheng Ni<sup>1</sup>, Peihong Ni<sup>1</sup>, and Suming Xiong<sup>1</sup>

<sup>1</sup>College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China

<sup>2</sup>Electronic and Computer Engineering Department, Duke University, Durham, NC 27708-0291 USA

<sup>3</sup>Faculty of Engineering and Information Technology, Griffith University, Brisbane QLD 4111, Australia

<sup>4</sup>Electrical Engineering Department, Hong Kong Polytechnic University, Hong Kong

**The support vector machine (SVM) is proposed as a response surface model to accelerate the solution speed of metaheuristic algorithms in solving inverse problems. The detail formulations of the SVM regression model using  $\varepsilon$ -insensitive loss function are derived. Primary numerical results are reported to demonstrate the feasibility, performance, and robustness of the proposed SVM based response surface model for solving both mathematical functions and engineering design problems.**

**Index Terms**—Inverse problem, metaheuristic algorithm, response surface model, support vector machine (SVM).

## I. INTRODUCTION

**M**ETAHEURISTIC algorithms have enjoyed great success in global optimization of electromagnetic devices. However, most metaheuristic algorithms require a large number of function evaluations before a well acceptable solution can be found. Generally, function evaluations in engineering design problems, especially in inverse problems, involve sophisticated and time consuming computer simulations such as finite-element analysis and hence such algorithms become computationally expensive. Therefore, computationally efficient approximation techniques are necessary. For this purpose, different response surface methodologies, such as the radial basis function, the moving least squares approximation, and the neural networks, have all been developed and employed with some success in solving typical electromagnetic design problems. Nevertheless, there is no universal approximation technique that is equally successful for all engineering problems. Therefore, it is essential to explore different approximation models.

Recently, a novel type of learning machine, called the support vector machine (SVM) [1], has received increasing attention in various fields, ranging from its original application in pattern recognition to other applications such as function regression because of its remarkable characteristics such as good generalization performance, absence of local minima, ability to model nonlinear relationships, and the sparse representation of a solution. Indeed, SVM has become the state-of-the-art technique for linear and nonlinear input–output knowledge discovery [2]. There are also a few applications of this new technique in the computational electromagnetics [3]. This paper reports the findings of a study that explores the potential of SVM as a response surface model (RSM) in applications of inverse problems to accelerate the solution speed of a metaheuristic algorithm.

## II. SVM REGRESSION AND ITS APPLICATION

### A. SVM Regression

The SVM developed by Vapnik and his coworkers [1] is a novel learning machine for classification and regression problems. It is based on the structural risk minimization (SRM) principle which seeks to minimize an upper bound of the generalization error that includes the sum of the training error and a confidence level dependent on the Vapnik–Chervonenkis (VC) dimension. This is different from the commonly used empirical risk minimization (ERM) principle that minimizes the training error only. Generally, SVM achieves higher generalization performance than traditional neural networks that implement ERM principle in solving many machine learning problems. Moreover, training SVM is essentially equivalent to finding the solution of a linearly constrained quadratic programming problem while ensuring the solution of the SVM is always unique and globally optimal.

Given a set of data points  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  with  $x \in R^d$  and  $y \in R$ , the approximation or regression of function  $f$  using SVMs is generalized as

$$f(x) = f(x, w) = \langle x, w \rangle + b \quad (1)$$

where  $\langle \bullet, \bullet \rangle$  denotes the canonical dot product of two vectors.

To determine the coefficients  $w$  and  $b$ , the following  $\varepsilon$ -insensitive loss function is defined in (2) and used because, unlike other ones such as the quadratic and Huber loss functions where all the data points are support vectors, the SV solution based on this loss function is sparse

$$L_\varepsilon[y, f(x, w)] = \begin{cases} 0 & \text{if } |y - f(x, w)| \leq \varepsilon \\ |y - f(x, w)| - \varepsilon & \text{otherwise} \end{cases} \quad (2)$$

Both parameters  $w$  and  $b$  are determined by minimizing the following cost function

$$\text{mimimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n L_\varepsilon(f(x_i), y_i) \quad (3)$$

where  $\|w\|^2/2$  is a term characterizing the model complexity (flatness of  $f(x)$  and  $L_\varepsilon$ ), and  $C$  is a regularization constant

Manuscript received October 07, 2008. Current version published February 19, 2009. Corresponding author: S. Yang (e-mail: shiyouyang@yahoo.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMAG.2009.2012690

determining the tradeoff between the model complexity of  $f(x)$  and the largest deviation  $\varepsilon$ , which is tolerated in the model [4].

To achieve a well-performed optimal model, the values of the two parameters  $C$  and  $\varepsilon$  have to be chosen carefully.

By means of some mathematical manipulations and introducing the slack variables  $\xi_i, \xi_i^*$  the “primal form” in the original data space of SVM regression is derived as

$$\begin{aligned} & \text{minimize} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ & \text{subject to} \quad \begin{cases} y_i - f(x_i, w) - b \leq \varepsilon + \xi_i^* \\ f(x_i, w) + b - y_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0. \end{cases} \end{aligned} \quad (4)$$

Since ill-posed problems are often encountered in the numerical implementation of the primal form, the dual form of (4) in the feature space is introduced and used, i.e.,

$$\begin{aligned} & \text{maximize} \quad -\frac{1}{2} \sum (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle \phi(x_i), \phi(x_j) \rangle \\ & \quad - \varepsilon \sum_{i=1}^n (\alpha_i - \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) \\ & \text{subject to} \quad \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, \quad 0 \leq \alpha_i, \alpha_i^* \leq C \end{aligned} \quad (5)$$

where  $\alpha_i$  and  $\alpha_i^*$  are Lagrange multipliers, the vector inner-product  $\langle \phi(x_i), \phi(x) \rangle$  represents the mapping from the input space to the feature space using a function  $\phi(x)$ .

The solution of the quadratic programming problem of (5) is always unique, and is obtained readily by using a quadratic programming optimizer. Consequently, (1) becomes

$$f(x) = \sum_{i=1}^m (\alpha_i - \alpha_i^*) \langle \phi(x_i), \phi(x) \rangle + b \quad (6)$$

Based on the Karush–Kuhn–Tucker conditions, only a (small) number of coefficient  $\alpha_i$  and  $\alpha_i^*$  will be nonzero, and the data points associated with these nonzero  $\alpha_i$  and  $\alpha_i^*$  refer to the support vectors of the model. Moreover, by means of the mapping of the inner-product  $\langle \phi(x_i), \phi(x) \rangle$  from the input data space to the feature space, a nonlinear regression in data space can be performed in the feature space in a linear way. Mathematically, an inner product in feature space has an equivalent kernel,  $K(x_i, x)$ , in input space. Consequently, the close-form approximation of (6) is changed by using a kernel function to become

$$f(x) = \sum_{i=1}^m (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (7)$$

where  $m$  is the number of support vectors.

### B. A Global Optimizer Based on SVM Regression

To reduce the number of function evaluations involving computationally expensive procedures such as finite-element analysis without sacrificing the computation accuracy, a fast global

optimizer based on the proposed SVM regression-based RSM and a metaheuristic algorithm is proposed and its iterative procedures can be described as follows.

---

#### Procedure A Fast Global Optimizer

---

Begin

Initialize a set of sample points; calculate the values of the objective/constraint functions at those sample points by using a computationally expensive approach;

Repeat

- 1) Reconstruct the optimal problem based on the SVM regression technique and solve it by using a tabu search method; Report all the searched local/global optimal solutions;
- 2) Solve the original optimal problem using a nonlinear simplex method starting from the newly searched local optimal solutions to find the “improved” ones;
- 3) Compare the local optimal solutions of (1) and (2); If a significant error exists for some solutions, intensify the sample points around the specified point and compute the values of the objective/constraint function and then go to (1); Otherwise, set Stopcriterion = “true”;

Until Stopcriterion=“true”.

End

## III. NUMERICAL EXPERIMENTS

To investigate the feasibility, performance, and robustness of the proposed SVM regression technique for function approximations and its application to computationally heavy inverse problems to accelerate the solution speed of a metaheuristic algorithm, extensive experiments are conducted on both standard mathematical functions and engineering problems. However, due to the space constraint which precludes an in-depth reporting of all numerical results in this paper, only the results on two typical case studies are presented.

### A. Mathematical Validation

To demonstrate the approximation power and robustness of the proposed SVM and to explore the influence of model parameters on its performances, a one-dimensional mathematical function is firstly approximated. This function is defined as

$$f(x) = 2 \cos x \quad (-1 \leq x \leq 1). \quad (8)$$

To implement the proposed SVM regression model, the Gaussian function is employed as the kernel function. The general form of the Gaussian kernel function reads as

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right). \quad (9)$$

TABLE I  
PERFORMANCE COMPARISON OF DIFFERENT APPROXIMATION TECHNIQUES FOR  
THE RECONSTRUCTION OF THE MATHEMATICAL FUNCTION

Method	Spar s i t y	Er r o r _m e t r i c
Pr o p o s e d SVM ( $\epsilon=0.1$ )	74/101	0.85
Pr o p o s e d SVM ( $\epsilon=0.5$ )	46/101	1.25
RBF	100%	0.64
MLS	70%	0.99

To evaluate the performances of a regression model, the following error metric is defined and used:

$$er r o _m e t r i c = \sum_{i=1}^{N_a} |f_i^{c l o s e} - f_i^{a p p r o}| \quad (10)$$

where  $f_i^{c l o s e}$  and  $f_i^{a p p r o}$  are, respectively, the close form and the approximated values of the test function at assessing point  $i$ ;  $N_a$  is the number of the total assessing points.

In the numerical experiments of the proposed SVM regression model, the parameter  $\epsilon$  is experimentally set to two different values, i.e., 0.1 and 0.5; the regularization parameter  $C$  is set to 10 000; the value  $\sigma$  of the Gaussian kernel function is set to 0.4 times the input range of the input data [5]. A total of 101 points are generated randomly in the feasible space to act as sample (training) points, and 201 equidistance points in the parameter space of  $[-1, 1]$  are used to assess the error metric of (10). To show the merits and demerits of the proposed SVM, this mathematical function is reconstructed by using the proposed SVM, a moving least squares (MLS) approximation, and a globally supported radial basis function (RBF) approximation technique, by using the same sample and assessing points. The globally supported RBF used in this case study is a multi-quadrics

$$h(r) = \sqrt{r^2 + 0.1}. \quad (11)$$

The MLS employed is the same as that described in [6]. Under such initial conditions and parameter settings, the error metrics and the sparsity of the coefficient matrix of the three approximated techniques are given in Table I. It should be pointed out that the sparsity in this table means the percentage of nonzero elements to the total number of elements of the coefficient matrix for the RBF and MLS, and the ratio of the number of support vectors to 101 of the total sample points for the proposed SVM. Moreover, to explore the influence of the choice of the data set points on the performance of the proposed SVM, this test function is independently reconstructed 100 times with different sample point sets generated randomly. The error metrics for these 100 independent runs of the proposed SVM are given in Fig. 1, where the error metrics are normalized with 0.85 being used as the base. From these numerical results one can deduce that

- 1) in terms of sparsity as a whole, the proposed SVM is comparable to MLS which is a well developed local approximation technique;
- 2) in terms of solution quality, the proposed one with the proper selection of model parameters is comparable to the widely recognized and entrusted globally supported RBF, and is slightly superior to the MLS;

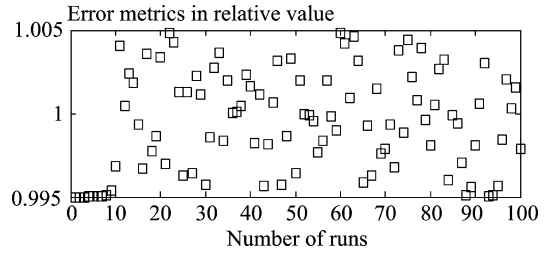


Fig. 1. Error metrics for 100 runs of the proposed SVM under different initial sample points sets in the construction of the mathematical function.

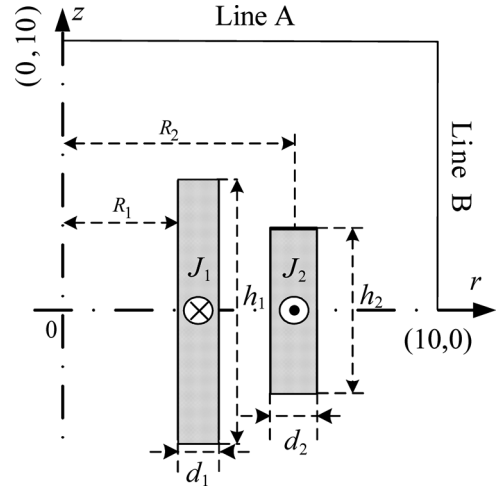


Fig. 2. Schematic diagram of the SMES.

- 3) even if a large deviation is used in the proposed SVM, the error metric is still acceptable while the support vectors take less than half of the total sample points;
- 4) The proposed SVM approximation technique is robust against differences in the distributions of sample points.

## B. Application

The Team Workshop problem 22 of a super-conducting magnetic energy storage (SMES) configuration with three free parameters [7] is solved as an example of the proposed SVM regression technique for engineering applications.

As shown in Fig. 2, the SMES configuration is optimized with respect to the following objectives. 1) The stored energy in the device should be 180 MJ. 2) The magnetic field must not violate a certain physical condition which guarantees superconductivity (quench condition of the superconductor). 3) The stray field (measured at a distance of 10 m from the device) should be as small as possible.

Consequently, this optimal design problem is mathematically formulated as

$$\begin{aligned} \text{minimize } f &= w_1 \frac{B_{\text{stary}}^2}{B_{\text{norm}}^2} + w_2 \frac{|Energy - E_{\text{ref}}|}{E_{\text{ref}}} \\ \text{subject to } J_i &\leq (-6.4 |(B_{\text{max}})_i| + 54) (A/\text{mm}^2) (i = 1, 2) \end{aligned} \quad (12)$$

where  $Energy$  is the stored energy in the SMES device,  $E_{\text{ref}} = 180$  MJ,  $B_{\text{norm}} = 2 \times 10^{-4}$  T,  $w_1$  and  $w_2$  are two weighting

TABLE II  
PERFORMANCE COMPARISONS OF DIFFERENT APPROXIMATION TECHNIQUES FOR SOLVING TEAM WORKSHOP PROBLEM 22

	$R_2$	$h_2/2$	$d_2$	$f_{opt}$	No. of FEM Computations
MLS	3.09	0.241	0.390	$8.21 \times 10^{-2}$	1000+52+41+45
RBF	3.10	0.240	0.388	$8.19 \times 10^{-2}$	1000+55+36
Proposed SVM	3.09	0.243	0.385	$8.18 \times 10^{-2}$	1000+45+56

TABLE III  
FINAL OPTIMAL RESULTS SEARCHED BY USING THE PROPOSED METHOD FOR THE FULL VERSION PROBLEM

Results	$R_1(m)$	$R_2(m)$	$h_1/2(m)$	$h_2/2(m)$	$d_1(m)$	$d_2(m)$
Proposed	1.5705	2.1020	0.7849	1.4201	0.6001	0.2574
By IGTE	1.5703	2.0999	0.7846	1.4184	0.5943	0.2562

TABLE IV  
FINAL OPTIMAL RESULTS SEARCHED BY USING THE PROPOSED METHOD FOR THE FULL VERSION PROBLEM

Results	$J_1(MA/m^2)$	$J_2(MA/m^2)$	$B_{stray}^2(T^2)$	Energy(MJ)	$f_{obj}$
Proposed	17.2470	-12.9643	$2.2005 \times 10^{-10}$	180.2001	$6.6219 \times 10^{-3}$
By IGTE	17.3367	-12.5738	$2.1913 \times 10^{-10}$	179.9924	$5.5203 \times 10^{-3}$

factors,  $B_{stray}^2$  is a measure of the stray fields evaluated along 22 equidistant points from Line A and Line B of Fig. 2 by

$$B_{stray}^2 = \sum_{i=1}^{22} (B_{stray})_i^2 / 22. \quad (13)$$

For performance comparison purposes, three different approximation techniques, i.e., the moving least squares approximation, a globally supported radial basis function, and the proposed SVM based ones, are utilized in the construction of the optimal problem in Step 2 of the fast global optimizer as described in Section II-B. The globally supported RBF used in this application is

$$h(r) = (r^2 + 0.1)^{0.04}. \quad (14)$$

The MLS used in this case study is the same as that used in the case of the mathematical function test. In the numerical implementation of the proposed SVM regression model, the same Gaussian function but with a different parameter as defined in (9) is used as the kernel function; and the values for parameters  $C$  and  $\varepsilon$  are set, respectively, to 1000 and 0.01.

In the numerical experiments of this case study for the three aforementioned approximation methods, a total of 1000 sampling points are first generated independently and identically in the feasible space. The objective and constraint functions on these sampling points are determined by using finite-element analysis, and the optimal problem is then reconstructed using the function values of these sample points and they are solved iteratively following the procedures as explained previously.

With these models and parameter settings, it is found that: 1) two iterative cycles are required for the proposed and the globally supported RBF approximation techniques, and three iterative cycles are needed for the MLS-based one, before the corresponding procedure converges to an acceptable solution; 2) 715 sampling points out of the total 1000 ones are identified as the support vectors for the proposed SVM model.

The detail comparison results of other performance parameters are tabulated in Table II. From these primary numerical results it is clear that:

- (1) in view of the approximation accuracy, the proposed SVM model is comparable to the globally supported RBF one since the same number of iterative cycles is required by both models;
- (2) in view of the solution efficiency, the proposed one is comparable to the compactly supported MLS model, since only about 70% sampling points are support vectors.

Moreover, the proposed methodology is also used to solved the full (eight variables) version of this problem to further validate its feasibility for complex engineering problems, and the final results are given in Tables III and IV. Again, these numerical results confirm positively the feasibility of the proposed methodology for solving complex inverse problems.

To sum up, these numerical experiments demonstrate that the salient characteristic of the proposed SVM models is that the efficiency and accuracy are well compromised compared with available response surface models.

#### IV. CONCLUSION

A support vector machine-based response surface model is proposed and used in the development of a fast global optimizer. The numerical results on two different applications demonstrate that the proposed approximation technique has the advantages of a globally supported RSM with high approximation accuracy, and also the merits of a well developed local RSM with sparsity of the coefficient matrix or high solution efficiency. For future work along this line, the authors are trying to give a theoretical exploration about the influence of the model parameters on the algorithm performances and the best strategy for the adaptive selection of the optimal model parameters in accordance to the characteristics of the reconstructed function.

#### ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 50777054 and the Specialized Research Fund for the Doctoral Program of Higher Education of China under Grant 20070335031.

#### REFERENCES

- [1] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
- [2] M. Lazaro, I. Santamaria, F. Perez-Cruz, and A. Artes-Rodriguez, "Support vector regression for simultaneous learning of a multivariate function and its derivatives," *Neurocomputing*, vol. 69, pp. 42–61, 2005.
- [3] Y. Yang, R. S. Chen, and Z. B. Ye, "Combination of particle-swarm optimization with least-squares support vector machine for FDTD time series forecasting," *Microw. Opt. Technol. Lett.*, vol. 48, pp. 141–144, 2006.
- [4] B. Ustun, W. J. Melssen, and L. M. C. Buydens, "Facilitating the application of support vector regression by using a universal Pearson VII function based kernel," *Chemomet. Intell. Lab. Syst.*, vol. 81, pp. 29–40, 2006.
- [5] V. Cherkarsky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural Netw.*, vol. 17, pp. 113–126, 2004.
- [6] S. L. Ho, S. Yang, P. H. Ni, and H. C. Wong, "Developments of an efficient global optimal design technique—a combined approach of MLS and SA algorithm," *COMPEL*, vol. 21, pp. 604–614, 2002.
- [7] TEAM Optimization Benchmark Problem 22. [Online]. Available: [www.igte.tugraz.at/archive/team/index.htm](http://www.igte.tugraz.at/archive/team/index.htm).