

WAVELET KERNEL SUPPORT VECTOR MACHINES FOR SPARSE APPROXIMATION¹

Tong Yubing Yang Dongkai Zhang Qishan

(Dept of Electronic Information Engineering, Beijing University of Aeronautics and astronautics,
Beijing 100083, China)

Abstract Wavelet, a powerful tool for signal processing, can be used to approximate the target function. For enhancing the sparse property of wavelet approximation, a new algorithm was proposed by using wavelet kernel Support Vector Machines (SVM), which can converge to minimum error with better sparsity. Here, wavelet functions would be firstly used to construct the admitted kernel for SVM according to Mercer theory; then new SVM with this kernel can be used to approximate the target function with better sparsity than wavelet approximation itself. The results obtained by our simulation experiment show the feasibility and validity of wavelet kernel support vector machines.

Key words Wavelet kernel function; Support Vector Machines (SVM); Sparse approximation; Quadratic Programming (QP)

I. Introduction

Sparse approximation is commonly a principle for signal decomposition. Let $f(\mathbf{x}, \boldsymbol{\alpha})$ be an approximation function of $f(\mathbf{x})$ with the form:

$$f(\mathbf{x}; \boldsymbol{\alpha}) = \sum_{i=1}^n \alpha_i \varphi_i(\mathbf{x}) \quad (1)$$

where $\varphi \equiv \{\varphi_i(\mathbf{x})\}_{i=1}^n$ is a fixed set of basis functions. If φ is not an orthogonal basis, it is possible that many different sets of coefficients will achieve the same error on a given data set. Sparse approximation looks, among all the approximation functions that achieve the same error, for the one with the smallest number of non-zero coefficients^[1]. Eq.(1) can be rewritten with the following form:

$$\mathbf{s} = \boldsymbol{\Phi} \boldsymbol{\alpha} \quad (2)$$

where $\boldsymbol{\Phi}$ is the basis matrix, \mathbf{s} is the original signal. In other word, sparse approximation is also equivalent to finding the smallest number of non-zero coefficients within vector $\boldsymbol{\alpha}$.

Sparse approximation can also be equivalent to the search for the optimal space structure, which can be used to describe the original signal or original space structure with very low cost. By this, it is more efficient for signal compression or signal Principle Component Analysis (PCA).

Ideally, it will lead to an NP problem to

minimize $\|\boldsymbol{\alpha}\|_{L_0}$ ^[2]. Dictionary and atom are commonly used in sparse approximation^[3]. Recently, there are several popular approaches to obtain solution to Eq.(2), such as Method Of Frame (MOF), Match Pursuit (MP), Best Orthogonal Basis (BOB), Basis Pursuit (BP), Support Vector Machines (SVM) and wavelet *et al.* MOF is not sparsity preserving^[2]. MP is only good for orthogonal dictionaries and is sub-optimal in terms of sparse property^[4]. When the signal is composed of a moderate number of highly non-orthogonal components, BOB may not deliver sparse representation^[5]. BP requires the solution of a convex non-quadratic optimization problem, which involves considerably more effort and sophistication than MP^[1]. SVM uses a device called kernel mapping to map the data in input space to high-dimensional features space in which the problem becomes linearly separable. SVM can be used to approximate function with good sparsity after ε -insensitive loss function is introduced^[6]. In some cases, only the near-optimal representations of the original signal can be gained in terms of sparse property by using wavelet or wavelet package. So SVM is introduced here to enhance the sparse property of wavelet approximation or wavelet package approximation. Zhang Li has also researched wavelet SVM, on which the convergent speed and precision instead of the sparse property of signal approximation is concentrated^[7]. This paper, the sparse property of signal approximation is mainly concerned and simulation experiment is made with Wavelet Kernel Support Vector Machines (WKSVM). Experiment results show WKSVM

¹ Manuscript received date: October 24, 2004; revised date: August 18, 2005.

Communication author: Tong Yubing, born in 1977, male, Ph.D. candidate. School of Electronics and Information Engineering, Beijing University of Aeronautics and Astronautics. No.37, Xueyuan Road, Haidian District, Beijing 100083, China. ybtongbuaa@buaa.edu.cn.

feasible and valid to solve the sparse approximation problem of wavelet.

In Section II, sparse property of wavelet approximation is analyzed. In Section III, a new method is provided based on WKSVM for solving sparse approximation problem. Simulation experiment and results analysis are performed in Section IV. Conclusions are presented in Section V.

II. Wavelet Approximation Analysis

Binary discrete wavelet transform is defined with the form

$$W_f(j, k) = (f(t), \varphi_{j,k}(t)) \quad (3)$$

where “ \cdot ” denotes dot product. And mother wavelet with the form

$$\varphi_{j,k}(t) = 2^{j/2} \varphi(2^j t - k) \quad (4)$$

Discrete wavelet can approximate the object function with high precision, but it is still not fine enough as far as sparse property is concerned. Wavelet approximation or wavelet package approximation can not get the optimal representation in some cases in terms of sparse property.

At the j -th level, there are 2^j wavelets of width $n/2^j$. The wavelets at this scale are all circulant shifts of each other, with the shift of $n/2^j$ samples. But some atoms that can be used to describe the nature of the original function can not be generated by dilation and translation, so the solution of wavelet decomposition may not be sparse enough. Chen, *et al.* analyzed this condition by using stationary wavelet dictionary^[3].

Commonly, wavelet package algorithm can be attributed to find best wavelet package base, which is not translation-invariant. Especially, if a signal comprises different kinds of high-energy structure and locates in the same frequency space but different time position, the best wavelet package base does not exist which satisfies the request of all these structure. Then the sparse property of signal decomposition becomes worse. Orthogonal basis is only a part of dictionaries, the demand that best package find an orthogonal basis prevents it from finding a highly sparse representation.

III. Wavelet Kernel Support Vector Machines

1. SVM sparse approximation

Given ε -insensitive loss function with the

form:

$$|y - f(\mathbf{x}; \mathbf{w})|_\varepsilon = \begin{cases} 0, & |y - f(\mathbf{x}; \mathbf{w})| \leq \varepsilon \\ |y - f(\mathbf{x}; \mathbf{w})| - \varepsilon, & \text{other} \end{cases} \quad (5)$$

$f(\mathbf{x}; \mathbf{w}) = \sum_{i=1}^n w_i \varphi_i(\mathbf{x})$, an approximation function of the original function $f(\mathbf{x})$ is defined to be sparse if the coefficients have been chosen to minimize the following lost function

$$E[\mathbf{w}] = \left\| f(\mathbf{x}) - \sum_{i=1}^n w_i \varphi_i(\mathbf{x}) \right\|_{L_2}^2 + \lambda \|\mathbf{w}\|_{L_1} \quad (6)$$

where $\varphi \equiv \{\varphi_i(\mathbf{x})\}_{i=1}^n$ is a fixed set of basis functions. Using SVM to minimize the $E[\mathbf{w}]$ can be finished by solving the following Quadratic Programming (QP) problem with optimal results:

$$\begin{aligned} \max W(\mathbf{a}; \mathbf{a}^*) &= \sum_{i=1}^N y_i (a_i - a_i^*) - \varepsilon \sum_{i=1}^N (a_i + a_i^*) \\ &\quad - \frac{1}{2} \sum_{i=1}^N (a_i - a_i^*) (a_j - a_j^*) K(x_i, x_j) \end{aligned}$$

$$\begin{cases} \sum_{i=1}^N a_i^* = \sum_{i=1}^N a_i \\ \text{s.t.} & 0 \leq a_i^* \leq 1 \\ & 0 \leq a_i \leq 1 \\ & a_i a_i^* = 0 \end{cases} \quad (7)$$

then the approximation function $f(\mathbf{x}; \mathbf{w})$ can be rewritten into the following form:

$$f(\mathbf{x}; \mathbf{w}) = \sum_{i=1}^N \beta_i K(\mathbf{x}, x_i) + b \quad (8)$$

where $\beta_i = a_i^* - a_i$, $i = 1, \dots, N$; \mathbf{a}, \mathbf{a}^* are non-negative coefficients, which are the optimal results of the above QP problem in Eq.(7), b is a constant and can be omitted for the admitted kernel function with constant component; $K(\cdot, \cdot)$ is a given function called as kernel function of SVM. Due to the nature of this QP problem, only a number of coefficients β_i will be different from zero and the input data point x_i associated to them are called support vectors. The number of support vectors reflects the sparse degree of signal approximation.

2. WKSVM

(1) Kernel function of SVM

The kernel function of SVM can be dot-product or translation-invariant^[7]. Mercer theory has given the conditions that admitted dot-product kernel must satisfy.

$K(u, v)$ can be an admitted kernel function

and be written into the following form:

$$K(u, v) = \sum_{k=1}^{\infty} \alpha_k \varphi_k(u) \varphi_k(v) \quad (9)$$

If for all $g(u) \neq 0$ and $\int g^2(u)du < \infty$, then

$$\iint K(u, v)g(u)g(v)dudv > 0 \quad (10)$$

(2) Wavelet kernel function and WKSVM

Let $h(x)$ be a mother wavelet, let d and t denote the dilation and translation factor, respectively. $d, t \in \mathbb{R}$, if $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^N$, then translation-invariant wavelet kernels that satisfy the translation-invariant kernel conditions are

$$K(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^N h\left(\frac{x_i - x'_i}{d}\right) \quad (11)$$

And the dot-product kernels are

$$K(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^N h\left(\frac{x_i - t_i}{d}\right) h\left(\frac{x'_i - t'_i}{d}\right) \quad (12)$$

The approximation function is

$$f(\mathbf{x}; \mathbf{x}') = \sum_{i=1}^N (a_i - a_i^*) K(\mathbf{x}, x'_i) + b \quad (13)$$

where a_i, a_i^* can be obtained by solving the QP problem of Eq.(7). Clearly, $K(\mathbf{x}, \mathbf{x}')$ in Eq.(12) or Eq.(13) is the new admitted wavelet kernel function for SVM.

IV. Simulation Experiment and Results Analysis

For wavelet approximation, several wavelet functions could be used to approximate original functions, such as Meyer wavelet, Haar wavelet, Daubechies wavelet, etc. For enhancing the sparsity of approximation, wavelet can be used to make tensor product and construct the admitted kernel function for SVM according to Mercy theory. Here, Daubechies wavelet is used in the simulation experiment to illustrate that WKSVM can effectively enhance the sparse property of wavelet approximation. Daubechies 4 (db4) and Daubechies Kernel SVM (dbKSVM) are used to approximate the same original function. Original function is $f(x)=\cos(\exp(x))$, $x \in (0, 3.5)$.

db4 wavelet is used to approximate the original function respectively with 107 non-zero wavelet coefficients in Fig.1 and 53 non-zero wavelet coefficients in Fig.2. And 107 and 53 support vectors are also respectively used to approximate the original function by dbKSVM in Fig.3 and Fig.4. The number of non-zero wavelet coefficients or support vectors reflects the sparse property of approximation methods. In each of

the figures, the middle sub-figure (b) describes the curve of the approximation function and sub-figure (c) clearly reflects the error between the approximate value and the target value.

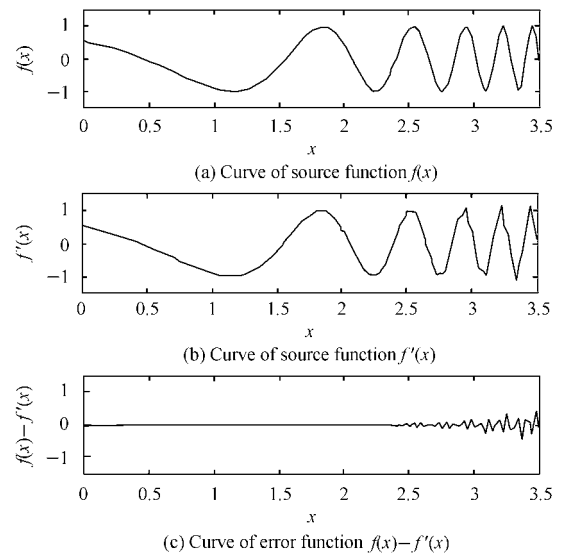


Fig.1 Approximation by db4 (107)

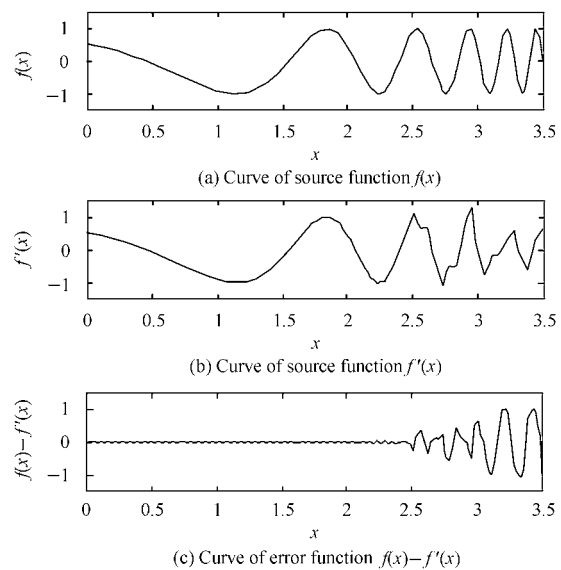


Fig.2 Approximation by db4 (53)

The more non-zero wavelet coefficients or support vectors are used, the smoother the approximation function curve and the error curve. But the difference between Fig.1 and Fig.2 is more obvious than that between Fig.3 and Fig.4; especially the difference of the approximation function curve in Fig.1 and Fig.2 is far more obvious than that in Fig.3 and Fig.4. That means, with the decrease of the number of non-zero wavelet coefficients used in wavelet approximation, the approximation function became worse quickly. But that is not the case for dbKSVM. It still keeps good approximation curve even though the number of support vectors is reduced.

In other word, under the admitted error bound, WKSVM can get results with far better sparse property than wavelet approximation. Comparing Fig.1 and Fig.3 or Fig.2 and Fig.4, at the bottom of the input space, there is visible error that is prone to increase by wavelet approximation. WKSVM approximation with ε -insensitive loss function could be equivalent to an QP problem with the optimal result, which made the waveform of the approximation function curve undulate but never deviate much from the original function. The standard deviation of the points on the error curve in every figure methods also demonstrates that.

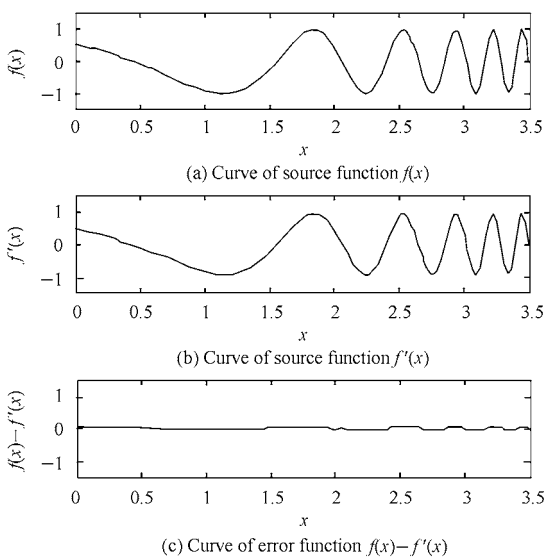


Fig.3 Approximation by dbKSVM (107)

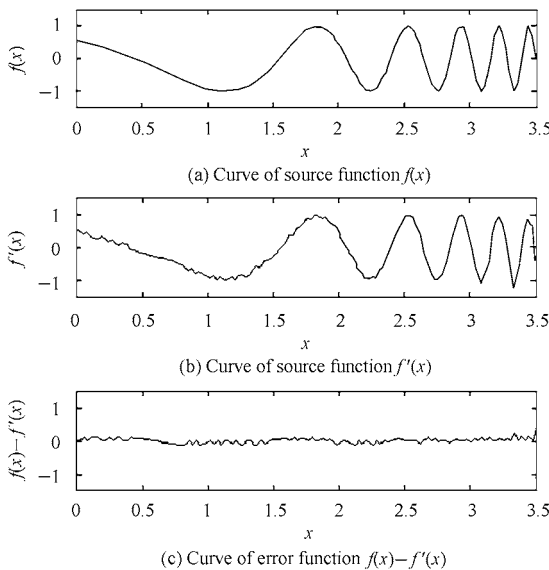


Fig.4 Approximation by dbKSVM (53)

V. Conclusions

Wavelet and SVM can be used together in pattern recognition, in which wavelet is used to

extract pattern characteristics and SVM is used to classify them^[8]. By contrast, wavelet and SVM are combined more closely in this paper. New SVM kernel function can be constructed by using wavelet, then wavelet kernel SVM is generated. In general, convergent speed, approximating precision and sparsity are all important and meaningful for signal approximation. More papers and experiments focused on the former. However this paper mainly discussed the later. The good sparse property of wavelet kernel SVM approximation is emphasized and tested here. The smooth property of wavelet approximation was gained at the sufficient cost of sparse property. But WKSVM was not. Combining SVM and wavelet, WKSVM can enhance the sparse property of wavelet. And the results of simulation experiment show the feasibility and validity of wavelet kernel SVM in sparse approximation.

Tab.1 Standard deviation

Figure	Standard deviation	Numbers of non-zero wavelet coefficients or support vectors
Fig.1	0.4846	107
Fig.2	6.1732	53
Fig.3	0.2858	107
Fig.4	0.7114	53

References

- [1] F. Girosi. An equivalence between sparse approximation and support vector machines. *Neural Computation*, **10**(1998)8, 1455–1480.
- [2] J. S. Kandola. Interpretable Modeling with Sparse Kernels. University of Southampton, 2001, 83–87.
- [3] S. S. B. Chen, D. L. Donoho, M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, **20**(1998)1, 33–61.
- [4] A. J. Smola, B. Scholköpfung. A tutorial on support vector regression. ESPRIT Working Group in Neural and Computational Learning II, NeuroCOLT2 27150, 1998, 48–50.
- [5] Stephane Mallet. A Wavelet Tour of Signal Processing. Beijing, China Machine Press, 2002, 286–330.
- [6] V. N. Vapnik. An overview of statistical learning theory. *IEEE Trans. on Neural Networks*, **10** (1999)5, 988–999.
- [7] Zhang Li, Zhou Weida, Jiao Licheng. Wavelet support vector machine. *IEEE Trans. on System, Man and Cybernetics-Part B: Cybernetics*, **34** (2004)1, 34–39.
- [8] Zhu Hailong. Face detection based on wavelet transform and support vector machine. *Journal of Xi'an Jiaotong University*, **36**(2002)9, 947–980, (in Chinese).