On the Relationship between LS-SVM, MSA, and LSA

Yatong Zhou, Taiyi Zhang, and Liejun Wang

Dept. of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an, 710049 P. R. China

Summary
A common task in signal processing is to approximate adequately a signal. It is crucial to understand various methods to this task. In this paper we survey three different approximation methods: least squares support vector machine (LS-SVM), multiresolution signal approximation (MSA), and least squares approximation (LSA) to highlight their mathematical relationship. Based on the theoretical analysis, we prove that LS-SVM is identical to the LSA with the minimum norm solution, and MSA is identical to the LSA with the least squares solution. Therefore, both LS-SVM and MSA can be derived as specific instances of LSA.

Key words:
Signal processing, support vector machine, multiresolution, approximation, least squares, relationship

1. Introduction

Signal approximation is of practical importance as it is closely related to the signal compression as well as many other signal processing techniques. It has been studied extensively over the past decades. Traditionally, approximation methods such as low-order polynomials approximation [1], piecewise linear approximation [2], spline approximation [3], kernel or projection based approximation [4], and least squares approximation (LSA) [5], *etc.*, have received considerable attention in the signal processing community. However, many new types of approximation methods, such as rough sets or neural network based approximation [6], support vector machines (SVM) [7], and MSA [8] have emerged recently that claim to have more flexibility and better approximation ability. Among these approximation methods developed, SVM and MSA are of the most successful methods.

The foundation of SVM has gained popularity due to its many attractive, analytic and computational features, and promising empirical performance [9]. It has been successfully extended from basic classification tasks to handle regression, operator inversion, density estimation and novelty detection. SVM used for approximation is introduced by Vapnik [7] and further investigated by many others. LS-SVM is a least square version for SVM that involves equality instead of inequality constraints and works with a least squares cost function [10]. This reformulation greatly simplifies the problem in such a way that the solution is characterized by a set of linear equations instead of a quadratic programming (QP). The equations can be efficiently solved by iterative methods such as conjugate gradient method [11].

Over the past few years, MSA has been successfully applied to approximate signals in many fields. It is formulated based on the study of wavelet analysis. MSA employs a coarse-to-fine strategy and provides a simple hierarchical approximation of the signals. At different resolutions level, the approximation of the signal characterizes different physical structures [8].

In this paper we look at approximation as the problem of recovering a signal f from given data set, being closest to the underlying signal $\breve{f}$. To express the concept of two signals being 'close' some measure criteria need to be defined. The most general criterion is the minimization of the mean squares error (MSE). LSA is a traditional and widely applicable approximation method [5]. It can be formulated in terms of minimization of the MSE that reflects the discrepancy between f and $\breve{f}$.

Given so many approximation methods, it has become difficult for an engineer to select the most appropriate method for the problems under study. So it would be very meaningful if one can know beforehand how much the solution of the different approximation methods is close to each other. Furthermore, It is valuable to investigate the connections between the traditional and currently developed approximation methods. The primary interest of this paper lies in describing the close mathematical relationship of above-mentioned three approximation methods: LS-SVM, MSA, and LSA. Based on the theoretical analysis, we prove that LS-SVM is identical to the LSA with the minimum norm solution, and MSA is equivalent to the LSA with the least squares solution. Therefore, both LS-SVM and MSA can be derived as specific instances of LSA.

The remainder of this paper is organized as follows. Sec.2 and Sec.3 are brief overviews of LS-SVM and MSA. Following that, in Sec.4 we review LSA. The relationship between LS-SVM, MSA, and LSA is proved in Sec.5. Finally we give some conclusions in Sec.6.

Manuscript received October 5, 2006.
Manuscript revised October 25, 2006.

2. Review on LS-SVM

Consider the following approximation problem (see Fig.1): let t_n ($n=1,2,\cdots,N$) be a set of N samples of a signal that contains the underlying signal $\breve{f}(x)$ and noise. Assume the signal $f(x)$ satisfactorily approximates signal $\breve{f}(x)$, we are asked to recover $f(x)$ based on the samples set $D=\{(x_n,t_n),n=1,\cdots,N\}$.

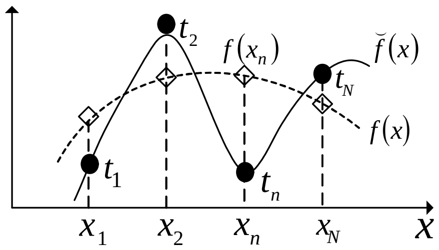

Fig.1 The approximation problem

LS-SVM has been developed for solving above problem. It allows the construction of an approximation by mapping the data set D implicitly into some feature space F through some mapping φ. Constructing a simple linear approximation in F then corresponds to a nonlinear approximation in sample space R. All can be done implicitly in F by using the kernel trick

$$K(x_n,x)=\varphi(x_n)\cdot\varphi(x).$$

LS-SVM aims at constructing an approximation of the form:

$$f(x)=\mathbf{w}^T\varphi(x)+b \quad , \tag{1}$$

where $\mathbf{w}$ is weight vector and b is bias term. To obtain $\mathbf{w}$ and b one solves the following constrained optimization problem

$$\min_{\mathbf{w},b,e}\frac{1}{2}\|\mathbf{w}\|^2+\frac{C}{2}\sum_{n=1}^{N}\xi_n^2 \quad , \tag{2}$$

subject to the equality constraints

$$t_n-\mathbf{w}^T\varphi(x_n)-b=\xi_n,\quad n=1,2,\cdots N \quad , \tag{3}$$

with ξ_n a error variable, C a regularization factor and ξ_n^2 the least squares cost function.

By constructing a Lagrange function from both the objective function and the corresponding constraints we yields the following expression

$$\min\ G(\boldsymbol{\alpha})=\frac{1}{2}\|\mathbf{w}\|^2+\frac{C}{2}\sum_{n=1}^{N}\xi_n^2 +\sum_{n=1}^{N}\alpha_n\left(t_n-\mathbf{w}^T\varphi(x_n)-b-\xi_n\right)$$

where α_n is Lagrange multipliers. The solution of $\boldsymbol{\alpha}$ and b is given by

$$\begin{bmatrix}0 & \mathbf{L}^T\\ \mathbf{L} & \boldsymbol{\chi}^T\boldsymbol{\chi}+C^{-1}\mathbf{I}\end{bmatrix}\begin{bmatrix}b\\ \boldsymbol{\alpha}\end{bmatrix}=\begin{bmatrix}0\\ \mathbf{T}\end{bmatrix}$$

where $\mathbf{T}=(t_1,t_{2,}\cdots t_N)^T$, $\boldsymbol{\chi}=(\varphi(x_1),\varphi(x_2),\cdots\varphi(x_N))$, $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_N)^T$, and $\mathbf{L}=(1,1,\cdots1)^T$. Finally, we substitute $\mathbf{w}$ into Eq. (1) and arrive at

$$f(x)=\sum_{n=1}^{N}\alpha_nK(x_n,x)+b \quad .$$

3. Multiresolution Signal Approximation (MSA)

Suppose there is a multi-resolution analysis in $L^2(R)$ such that the scale subspaces V_j and the wavelet subspaces W_j satisfy

$$V_{j-1}=V_j\oplus W_j \quad ,$$

$$\bigcup_{j\in Z}V_j=L^2(R) \quad ,$$

$$\bigcap_{j\in Z}V_j=\langle 0\rangle \quad ,$$

The symbol $\oplus$ denotes direct sum and Z is the set of integers. An orthogonal compactly supported wavelet basis of W_j is formed by the dilation and translation of a ψ, called the mother wavelet function and is given by

$$\psi_{jk}(x)=2^{-j/2}\psi\left(2^{-j}x-k\right).$$

Similarly, the orthogonal basis of V_j is given by

$$\phi_{jk}(x)=2^{-j/2}\phi\left(2^{-j}x-k\right)$$

where ϕ is called scaling function. The wavelet and scaling function satisfy the following dilation equation

$$\phi\left(2^{-j}x\right) = \sqrt{2}\sum_k h_{0k}\phi\left(2^{-j+1}x - k\right)$$

$$\psi\left(2^{-j}x\right) = \sqrt{2}\sum_k h_{1k}\psi\left(2^{-j+1}x - k\right)$$

The projection of a signal $\breve{f}(x)$ onto V_j has the form

$$P_j \breve{f}(x) = \sum_k c_k^{(j)}\phi_{jk}(x).$$

It can be seen as a smooth approximation of $\breve{f}(x)$ at resolution level j. Hence above equation can also be written as

$$f_j(x) = P_j \breve{f}(x) = \sum_k c_k^{(j)}\phi_{jk}(x) \quad . \tag{4}$$

The coefficient $c_k^{(j)}$ is the projection of $\breve{f}(x)$ on the basis function $\phi_{jk}(x)$; that is

$$\begin{aligned} c_k^{(j)} &= \int \breve{f}(x)\phi_{jk}(x)dx \\ &\approx \sum_{n=1}^{N} \breve{f}(x_n)\phi_{jk}(x_n)\Delta x \end{aligned} \quad . \tag{5}$$

where Δx is sample rate. In practice Eq. (5) has to be rewritten as

$$c_k^{(j)} \approx \sum_{n=1}^{N} t_n\phi_{jk}(x_n)\Delta x \quad ,$$

since $\breve{f}(x_n)$ is corrupted by noise thus only corrupted value t_n could be obtained.

4. Least Squares Approximation (LSA)

When approximating an unknown signal $\breve{f}(x)$, the signal is usually represented by a weighted sum of the linearly independent basis functions. The choice of the basis functions may vary depending on the application. Examples of basis functions are splines, complex exponentials, Gaussian radial basis function, or wavelets that have the best time frequency resolution. LSA is one of approximation method that represents an underlying signal $\breve{f}(x)$ using the basis functions $\{P_k(x)\}_{k=1}^{q}$, that is

$$f(x) = \sum_{k=1}^{q} u_k P_k(x) \quad .$$

To estimate u_k one can minimize the mean squares error, that is

$$\min \int e^2(x)dx = \min \int \left(f(x) - \breve{f}(x)\right)^2 dx \quad . \tag{6}$$

Usually the set $\{P_k(x)\}_{k=1}^{q}$ is complete in $L^2(R)$, which implies that Eq. (6) can be made arbitrarily small by selecting the order of the approximation q properly.

Eq. (6) takes a discrete form as following

$$\min_{u_k} \sum_{n=1}^{N}\left[\sum_{k=1}^{q} u_k P_k(x_n) - \breve{f}(x_n)\right]^2 \cdot \Delta x \quad , \tag{7}$$

since we need to be able to use it on the underlying signal $\breve{f}(x)$, which is only specified by the values it takes at the discrete set of points $\{x_n\}_{n=1}^{N}$. Take into account that deserting Δx has no influence on estimating u_k, and $\breve{f}(x_n)$ is corrupted by noise thus only corrupted value t_n could be obtained, Eq. (7) can be rewritten as

$$\min_{u_k} \sum_{n=1}^{N}\left[\sum_{k=1}^{q} u_k P_k(x_n) - t_n\right]^2 \quad . \tag{8}$$

5. Relationship between LS-SVM, MSA, and LSA

Let $\mathbf{A}$ be a $N \times q$ matrix with the element $a_{nk} = P_k(x_n)$, $\mathbf{t}_N = (t_1, t_2, \cdots, t_N)^T$ and $\mathbf{U} = (u_1, u_2, \cdots u_q)^T$ are two vectors. Arranging Eq. (8) in matrix form we have the following constrained optimization problem

$$\min_{\mathbf{E}} \mathbf{E}^T\mathbf{E} \quad , \tag{9}$$

subject to

$$\mathbf{AU} - \mathbf{t}_N = \mathbf{E} \quad . \tag{10}$$

The above optimization program exists two kinds of solutions by adding different constrains:

(i) The minimum norm solution

$$\mathbf{U} = \mathbf{A}^T \left(\mathbf{A}\mathbf{A}^T\right)^{-1} \mathbf{t}_N ,$$

obtained by adding the constrain $\min\|\mathbf{U}\|^2$, when $\mathbf{AU} = \mathbf{t}_N$ is an under-determined system.

(ii) The least squares solution

$$\mathbf{U} = \left(\mathbf{A}^T \mathbf{A}\right)^{-1} \mathbf{A}^T \mathbf{t}_N ,$$

obtained by adding the constrain $\min \mathbf{E}^T\mathbf{E}$, when $\mathbf{AU} = \mathbf{t}_N$ is an over-determined system.

Now we firstly discuss the relationship between LS-SVM and LSA. To eliminate bias term b in Eq. (3) we implement coordinate transformation in the feature space F [12]. Consequently, Eq. (3) is reformulated as

$$t_n - \mathbf{w}^T \mathbf{z}_n = \xi_n , \quad n = 1,2,\cdots N \quad , \tag{11}$$

where $\mathbf{z}_n = \left(z_{n1}, z_{n2}, \cdots z_{np}\right)^T$ is the transformation vector of $\varphi(x_n)$ with the dimension p. For simplicity we assume C be 1, thus Eq. (2) can be rewritten as

$$\min_{\mathbf{w},\xi} \frac{1}{2}\left(\|\mathbf{w}\|^2 + \|\boldsymbol{\xi}\|^2\right) \quad , \tag{12}$$

with $\boldsymbol{\xi} = \left(\xi_1, \xi_2, \cdots, \xi_N\right)^T$.

We suppose that $\mathbf{Z} = \left(\mathbf{z}_1, \mathbf{z}_2, \cdots \mathbf{z}_n\right)^T$, $\mathbf{A}_\mathbf{s} = \left(\mathbf{Z}, \mathbf{I}\right)$, $\mathbf{U}_\mathbf{s} = \left(\mathbf{w}, \boldsymbol{\xi}\right)^T$, and $\mathbf{I}$ is a unit matrix with rank p. Then Eq. (12) and (11) have the following equivalent expression

$$\min \frac{1}{2}\|\mathbf{U}_\mathbf{s}\|^2 \quad , \tag{13}$$

subject to

$$\mathbf{A}_\mathbf{s}\mathbf{U}_\mathbf{s} = \mathbf{t}_N \quad . \tag{14}$$

It is notable that Eq. (14) has the same form as Eq. (10) except the term $\mathbf{E}$. Furthermore, Eq. (14) is an under-determined system since $\mathbf{A}_\mathbf{s}$ is a $N \times (N+p)$ matrix. Motivated by Eq. (9) and (10), we can obtain a minimum norm solution to Eq. (14) by adding constrain (13). Therefore, it is quite logical to conclude that LS-SVM is identical to LSA with the minimum norm solution.

Similarly we can discuss the relationship between MSA and LSA. It has pointed out that MSA is an optimal approximation in the sense of minimal mean square error [8]. Hence the coefficient c_k^j obtained by Eq. (5) makes the mean square error reaching a minimum, i.e.

$$\sum_{n=1}^{N}\left[\sum_k c_k^{(j)} \phi_{jk}(n) - t_n\right]^2 = \min_{c_k^{(j)*}} \sum_{n=1}^{N}\left[\sum_k c_k^{(j)*} \phi_{jk}(n) - t_n\right]^2 \quad . \tag{15}$$

Given the length of sequence c_k^j be K, we have $K < N$ since there is an extraction process in Eq. (5) when the condition $j > 0$ is fulfilled. Now let $\mathbf{A}_\mathbf{r}^{(\mathbf{j})}$ be a $N \times K$ matrix with the element $a_{nk}^{(j)} = \phi_{jk}(n)$ and $\mathbf{U}_\mathbf{r}^{(\mathbf{j})} = \left(c_1^{(j)}, c_2^{(j)}, \cdots, c_K^{(j)}\right)^T$ be a $K \times 1$ vector, Eq. (15) can be arranged in the following constrained optimization problem

$$\min \mathbf{E}_\mathbf{r}^{(\mathbf{j})T} \mathbf{E}_\mathbf{r}^{(\mathbf{j})} \quad , \tag{16}$$

subject to

$$\mathbf{A}_\mathbf{r}^{(\mathbf{j})} \mathbf{U}_\mathbf{r}^{(\mathbf{j})} - \mathbf{t}_N = \mathbf{E}_\mathbf{r}^{(\mathbf{j})} \quad .$$

A quick inspection of Eq. (16) shows it has the same form as Eq. (9). Furthermore, there exist a least squares solution to Eq. (16) since we can infer that $\mathbf{A}_\mathbf{r}^{(\mathbf{j})} \mathbf{U}_\mathbf{r}^{(\mathbf{j})} = \mathbf{t}_N$ is an over-determined system from the fact $\mathbf{A}_\mathbf{r}^{(\mathbf{j})}$ is a $N \times K$ matrix. It is nature to conclude that MSA is identical to LSA with the least squares solution.

6. Conclusion

In this paper we survey two currently developed approximation method——LS-SVM and MSA, and one traditional method——LSA. As described in above sections, it is not surprising that close mathematical relationship between LS-SVM, MSA, and LSA exist. We prove that LS-SVM is identical to the LSA with the minimum norm solution, and MSA is identical to the LSA with the least squares solution. Therefore, both LS-SVM and MSA can be derived as specific instances of LSA.

Acknowledgment

The authors would like to express their cordial thanks to Dr. Xiaohe Li for his valuable advice.

References

[1] D. Natale, G. S. Desoli, D. D. Giusto, and G. Vernazza, "Polynomial approximation and vector quantization: a region-based integration," IEEE Trans. Communications, vol.43, pp.198-206, 1995.

[2] T. B. Nguyen and B. J. Oommen, "Moment-preserving piecewise linear approximations of signals and images," IEEE Trans. PAMI, vol.19, pp.84-91, 1997.

[3] Y. V. Zakharov and T. C. Tozer. Local spline approximation of time-varying channel model. Electronics Letters, vol.37, pp. 1408-1409, 2001.

[4] D. Donoho and I. Johnstone, "Projection-based approximation and a duality with kernel methods," Ann. Statist., vol. 17, pp. 58-106, 1989.

[5] S. Celebi and J. C. Principe, "Parametric least squares approximation using Gamma bases," IEEE Trans. Signal Processing, vol.43, pp.781-784, 1995.

[6] D. Wedge, D. Ingram, and C. Mingham, "On global local artificial neural networks for function approximation," IEEE Trans. Neural Networks, vol.17, pp.942-952, 2006.

[7] V. Vapnik, S. Golowich, and A. J. Smola, "Support vector method for function approximation, regression estimation, and signal processing," In: M. J. Jordan and S. A. Solla (eds.): Neural information processing system (NIPS), MIT press, Cambridge MA, pp.322-327, 1997.

[8] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," IEEE Trans. PAMI, vol.11, pp.674-693, 1989.

[9] V. Vapnik, The nature of statistical learning theory, Springer Verlag, New York, 1998.

[10] J. A. K. Suykens and J. Vandewalle, "Recurrent least squares support vector machines," IEEE Trans. Circuits and System, vol. 47, pp.1109-1114, 2000.

[11] J. A. K. Suykens, L. Lukas, and P. Dooren, "Least squares support vector machine classifiers: a large scale algorithm," In: S. Golowich and A. J. Smola (eds.): European Conference on Circuit Theory and Design (ECCTD), Stresa, Italy, pp.839-842, 1999.

[12] H. Yan, X. G. Zhang, and Y. D. Li, "Relation between a support vector machine and the least square method," J. Tsinghua Univ. (Sci. & Tech.), vol.41, pp.77-80, 2001 (in Chinese).

Yatong Zhou received the B.E. and M.E. degrees, from Chengdu Univ. of Technology in 1995 and 1999, respectively. He is currently pursuing the Ph. D. degree in the Dept. of Information and Information Engineering, Xi'an Jiaotong University, China. He is author or co-author of 18 publications. His current research interests include machine learning, pattern recognition, and adaptive signal processing. He is a student member of IEICE.

Taiyi Zhang received the B.E. and M.E. degrees, from Xi'an Jiaotong University in 1967 and 1981, respectively. He went on to work at Xi'an Optics and Precision Mechanics, Academia Sinica, where he was a senior researcher on image processing and communication system until 1998. He started work as a professor in mobile communication and WLAN at Xi'an Jiaotong University in the spring of 1998. He is author or co-author of 106 publications. His current research interests include digital mobile communications, pattern recognition, and image processing.

Liejun Wang received the B.E. and M.E. degrees, from Xinjiang Univ. in 1994 and 2003, respectively. He is currently pursuing the Ph. D. degree in the Dept. of Information and Information Engineering, Xi'an Jiaotong University, China. He is author or co-author of 8 publications. His current research interests include digital wireless communications, OFDM systems, and signal detection.