Projection wavelet weighted twin support vector regression for OFDM system channel estimation

Lidong Wang¹ · Yimei Ma¹ · Xudong Chang¹ · Chuang Gao¹ · Qiang Qu¹ · Xuebo Chen¹

Published online: 10 June 2020
© The Author(s) 2020

Abstract

In this paper, an efficient projection wavelet weighted twin support vector regression (PWWTSVR) based orthogonal frequency division multiplexing system (OFDM) system channel estimation algorithm is proposed. Most Channel estimation algorithms for OFDM systems are based on the linear assumption of channel model. In the proposed algorithm, the OFDM system channel is consumed to be nonlinear and fading in both time and frequency domains. The PWWTSVR utilizes pilot signals to estimate response of nonlinear wireless channel, which is the main work area of SVR. Projection axis in optimal objective function of PWWTSVR is sought to minimize the variance of the projected points due to the utilization of a priori information of training data. Different from traditional support vector regression algorithm, training samples in different positions in the proposed PWWTSVR model are given different penalty weights determined by the wavelet transform. The weights are applied to both the quadratic empirical risk term and the first-degree empirical risk term to reduce the influence of outliers. The final regressor can avoid the overfitting problem to a certain extent and yield great generalization ability for channel estimation. The results of numerical experiments show that the proposed algorithm has better performance compared to the conventional pilot-aided channel estimation methods.

Keywords Twin support vector regression · Wavelet transform · Unconstrained minimization · Channel estimation · OFDM

1 Introduction

Recently, with the rapid development of wireless communication technology, the demand for high-speed data transmission has increased rapidly. Broadband communication is an effective technology to provide high data rate transmission. However, the increase of bandwidth will result in less sampling period than channel delay spread, especially in multipath scenarios, which gives rise to frequency-selective channels. Carrier frequency-offset due to the oscillators' mismatch, together with high relative mobility between the transmitter

✉ Lidong Wang
wangld5600@163.com

Extended author information available on the last page of the article

and receiver cause the transmission channel to change rapidly in time, which is referred to as the time-selectivity of the channel. The channel with both the frequency-selectivity and the time-selectivity is called doubly-selective in wireless communications. Orthogonal frequency division multiplexing (OFDM) is an attractive multi-carrier modulation scheme, which divides the whole bandwidth into multiple overlapping narrowband subchannels to reduce the symbol rate. OFDM has the advantages of high spectral efficiency and simple structure of single tap equalizer. The bandwidth of each subcarrier in OFDM systems is small enough that each subcarrier is considered to experience flat fading in the frequency-selective channel. The narrow band nature of subcarriers makes the signal robust against the frequency-selectivity, and the inter-symbol interference (ISI) can be easily eliminated by inserting a cyclic prefix (CP) in front of each transmitted OFDM block. However, OFDM is relatively sensitive to the time-selectivity of the mobile channel. Time variation of the channel within an OFDM symbol duration results in intercarrier interference (ICI) (Wu and Fan 2016) and lead to an irreducible error floor in conventional receivers, which further degrades the performance. OFDM has strong robustness in high delay spread environment, which can eliminate the need for equalizing delay spread effect. This feature allows for higher data rates and leads to the selection of OFDM as the standard for digital audio broadcasting, digital video broadcasting, some wireless local area networks, long-term evolution and the next generation cellular systems. It has been proposed to be adopted in high speed train broadband communication system (Yu et al. 2016). The performance of the OFDM systems is affected by channel estimation, timing synchronization and mobility. The ICI may become more severe as mobile speed, carrier frequency or OFDM symbol duration increases (Sheng et al. 2017). OFDM can be implemented by using a coherent or non-coherent detection technique. Coherent detection methods usually provide higher signal-to-noise ratio gain than incoherent methods because they use channel state information. This implies, however, the receiver is more complex because channel state information is usually obtained through channel estimation. In order to achieve acceptable reception quality for applications with high latency and Doppler spread, it is essential to optimize the design of channel estimators.

In highly selective multipath fading channels, the channel response presents complex nonlinearity. If linear method is used, the estimation accuracy will be reduced (Charrada and Samet 2016). Therefore, it is necessary to use the nonlinear channel estimation method to improve the estimation performance. While support vector regression (SVR) developed from support vector machine (SVM) is suitable for regression of nonlinear systems. The SVM is proposed based on the principle of statistical learning theory and Vapnik–Chervonenkis (VC) dimensional theory, which is a promising machine learning approach that has been adopted in classification and regression (Vapnik 1995). The adoption of the kernel trick can be used to extend applications to nonlinear situations. It has become one of the most powerful tools for pattern classification and regression (Xu et al. 2017). So far, some SVR algorithms have been used in wireless channel estimation. Matilde et al. (2004) developed a multiple-input multiple-output channel estimation method based on SVR, but the channel was assumed to be non-selective. Djouama et al. (2014) and Charrada and Samet (2016) proposed OFDM channel estimation method based on SVR for different application scenarios. However, the methods mentioned above are based on the basic SVR, there are still shortcomings in computational complexity and performance. Along with continuous research progress, many modifications have been proposed in recent years. Peng (2010) proposed a twin support vector regression (TSVR), which can increase the computational speed by solving two small-size quadratic programming problems (QPPs). Xu et al. (2017) proposed asymmetric ν -twin support vector regression,

which is a kind of twin SVR suitable for dealing with asymmetric noise. Peng et al. (2014) introduced a pair of projection terms in the optimization problem, which has the advantage of embedding the structural information of the data into the learning process, resulting in the reduction of empirical variance. Melki et al. (2017) studied multi-target regression and presented several models for problems with multiple outputs. Anand et al. (2019), Anagha et al. (2018) and Gupta and Gupta (2019) proposed the improved SVR algorithms based on the pinball loss function, Balasundaram et al. (2014) (2016a, Knowl. Inf. Syst.) studied the problems of Lagrangian SVR, and Anand et al. (2018) proposed a generalized ε -loss function for regression.

However, all of the training samples in most methods are considered to have the same status and are given the same penalties, which may degrade performance due to the influence of noise or outliers. It is useful to give the training samples different weights depending on their importance. Although Xu and Wang (2014) proposed K-nearest neighbor (KNN)-based weighted twin support vector regression, which uses local information of data to improve prediction accuracy, KNN-based methods are suitable for clustering sample regression, but not for time series such as channel estimation. Recently, an efficient projection wavelet weighted twin support vector regression (PWWTSSVR) was proposed in our work (Wang et al. 2019), which introduces a weight matrix based on wavelet transform and suitable for dealing with time-series data. PWWTSSVR is with good normalization performance and makes full use of data structure information. It is suitable for regression of nonlinear systems with sequence information as training samples. The application of the proposed algorithm in wireless channel parameter prediction, a system with strong nonlinearity, will improve the estimation performance. The idea of combining the TSVR and weights based on wavelet transform was also applied to the regression of steelmaking model (Gao et al. 2019).

Taking the merits of PWWTSSVR, the proposed channel estimation algorithm can be with good performance. The contributions of this paper are summarized as follows.

1. The improved TSVR is adopted for the first time to estimate doubly selective wireless channel parameters in OFDM system. This method solves the problem that the performance of most traditional estimation methods is degraded by linear assumptions.
2. Aiming at nonlinearity and the characteristics of time series of fading channels of OFDM system, we propose an improved TSVR, PWWTSSVR algorithm based on wavelet transform. It can be said that the method of calculating weight matrix by wavelet transform in this work is a new preprocessing angle. Wavelet transform a kind of time-frequency representation for signals, therefore the proposed method based on the wavelet theory is suitable for dealing with time series samples such as channel parameters. Additionally, compared with TSVR, the PWWTSSVR improves the regression performance by adding projection terms to the objective function.
3. In order to make better use of information of the received pilot signal and improve the channel estimation accuracy, weight matrix is introduced into the objective functions for channel parameters regression. The weights are inserted into both quadratic and first-degree terms to reduce the influence of outliers, which is likely to appear in the received pilot signal polluted by noise. Intuitively, the PWWTSSVR can make full use of the prior information of training channel response. The weight matrix and weight vector, which represent the distance of noised samples and its 'real position', can reflect the prior information of the training channel response. The larger weight is given to samples with smaller noise and the smaller weight is given to samples with larger noise.

This paper is organized as follows: Sect. 2 briefly describes OFDM system. Section 3 proposes projection wavelet weighted TSVR channel estimation. Experimental results are described in Sect. 4 to investigate the validity of our proposed algorithm, and Sect. 5 ends the paper with concluding remarks.

2 System model

Consider an OFDM system with N subcarriers experiencing doubly selective fading channel. The sequence to be transmitted $X(k)$ from QPSK or QAM constellation is parsed into blocks of N symbols and then transformed into a time-domain sequence using an N -point IFFT. To avoid inter-block interference (IBI), a cyclic prefix (CP) of length M that is equal to or larger than the channel order L (channel discrete multipath number is $L + 1$) is inserted at the head of each block. The time-domain signal $x(n)$ can be serially transmitted over the fading channel. $x(n)$ can be expressed as

$$x(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X(k) e^{j2\pi nk/N} \quad (1)$$

where $n = -M, \dots, N-1$, $k = 0, 1, \dots, N-1$. After CP is removed at the receiver, the received signal in time domain $y(n)$ can be expressed as

$$y(n) = \sum_{l=0}^L h_l(n)x(n-l) + \eta(n) \quad (2)$$

where $\eta(n)$ is additive white Gaussian noise (AWGN) with zero-mean, variance σ_n^2 and independent with each other, i.e. $E(\eta(n)\eta(m)) = \delta(n-m)\sigma_n^2$. $h_l(n)$ is the baseband-equivalent doubly selective channel impulse response of the l th path ($l = 0, 1, \dots, L$) at time n , which includes the physical channel as well as filters at the transmitter and receiver.

The matrix form of (2) can be expressed as

$$\mathbf{y} = \mathbf{h}\mathbf{F}^H\mathbf{X} + \boldsymbol{\eta} = \mathbf{h}\mathbf{x} + \boldsymbol{\eta} \quad (3)$$

where $\mathbf{X} = [X(0), X(1), \dots, X(N-1)]^T$, $\mathbf{x} = [x(0), x(1), \dots, x(N-1)]^T$, $\mathbf{y} = [y(0), y(1), \dots, y(N-1)]^T$, $\boldsymbol{\eta} = [\eta(0), \eta(1), \dots, \eta(N-1)]^T$, $(\cdot)^T$ denotes the transpose operation; $\mathbf{X}, \mathbf{x}, \mathbf{y}, \boldsymbol{\eta} \in \mathbb{C}^N$, and the channel matrix $\mathbf{h} \in \mathbb{C}^{N \times N}$ can be expressed as

$$\mathbf{h} = \begin{bmatrix} h_0(0) & \cdots & h_L(0) & \cdots & h_1(0) \\ \vdots & \ddots & & \ddots & \vdots \\ \vdots & & & \mathbf{0} & h_L(L-1) \\ h_L(L+1) & \cdots & h_0(L+1) & & \mathbf{0} \\ & \ddots & & \ddots & \\ \mathbf{0} & & h_L(N-1) & \cdots & h_0(N-1) \end{bmatrix} \quad (4)$$

where $\mathbf{F}^H$ is an N -point inverse discrete Fourier transform (IDFT) matrix, $(\cdot)^H$ denotes conjugate transpose and the entry of which $[\mathbf{F}^H]_{n,k} = (1/\sqrt{N}) \exp(j2\pi nk/N)$. Also define Fourier transform (DFT) matrix $\mathbf{F}$, and $[\mathbf{F}]_{n,k} = (1/\sqrt{N}) \exp(-j2\pi nk/N)$.

Perform Fourier transform on both sides of (2) and the following equations can be obtained,

$$\begin{aligned}\mathbf{F}\mathbf{y} &= \mathbf{F}\mathbf{h}\mathbf{F}^H\mathbf{X} + \mathbf{F}\boldsymbol{\eta} \\ \mathbf{Y} &= \mathbf{H}\mathbf{X} + \mathbf{F}\boldsymbol{\eta}\end{aligned}\quad (5)$$

where $\mathbf{Y} = [Y(0), Y(1), \dots, Y(N-1)]^T \in \mathbb{C}^N$ is received signal vector in frequency domain. $\mathbf{H} = \mathbf{F}\mathbf{h}\mathbf{F}^H \in \mathbb{C}^{N \times N}$ is a frequency domain channel matrix with inter carrier interference (ICI) induced by time variations of the channel, and the elements of which can be described as

$$\mathbf{H}(p, q) = \frac{1}{N} \sum_{l=0}^L e^{-j2\pi pl/N} \sum_{n=0}^{N-1} h_l(n) e^{-j2\pi(p-q)n/N} \quad (6)$$

where $p, q = 0, 1, \dots, N-1$. The off-diagonal elements of $\mathbf{H}$ is the ICI response. $\mathbf{H}$ can be divided into two parts, one part $\mathbf{H}_d \in \mathbb{C}^{N \times N}$ is to retain only the principal diagonal elements, the other one $\mathbf{H}_n \in \mathbb{C}^{N \times N}$ is to retain only the non-diagonal elements. Then (5) can be expressed as

$$\begin{aligned}\mathbf{Y} &= \mathbf{H}_d\mathbf{X} + \mathbf{H}_n\mathbf{X} + \mathbf{F}\boldsymbol{\eta} \\ &= \text{diag}(\mathbf{X})\mathbf{H}'_d + \mathbf{H}_n\mathbf{X} + \mathbf{F}\boldsymbol{\eta}\end{aligned}\quad (7)$$

$\mathbf{H}_n\mathbf{X}$ is the ICI component, $\text{diag}(\cdot)$ denotes the diagonal operators and $\mathbf{H}'_d \in \mathbb{C}^N$ is a column vector, the element of which is taken from the principal diagonal element of $\mathbf{H}_d$.

Because of Doppler effect, a subcarrier will be interfered by ICI of adjacent subcarriers. Figure 1 shows the simulation results of the ICI power influence of adjacent subcarriers on one subcarrier. The horizontal axis is the distance between adjacent subcarriers and a certain subcarrier, and the vertical axis is the average normalized power of ICI. In this simulation, subcarrier number of an OFDM $N = 64$, carrier frequency $f_c = 2.15$ GHz, mobile speed is 10/120/350 (km/h). As can be seen from Fig. 1 that most energy is distributed over the diagonal and its neighbors, the main ICI power is generated by adjacent subcarriers. The closer the subcarriers are, the greater the interference

Fig. 1 ICI power influence of adjacent subcarriers

to a certain subcarrier, and the greater the moving speed is, the more serious the influence is.

3 Projection wavelet weighted TSVR channel estimation

Linear interpolation methods, which are often adopted in typical channel estimation are inadequate for time-varying fading channel estimation, while TSVR is suitable for regression of nonlinear systems due to its kernel mapping technique, therefore we use this method for nonlinear channel estimation. In 2010, the TSVR was proposed by Peng (2010), which is an extension of the classification tool support vector machine to regression applications; the target of the regression problem is to acquire the relationship between inputs and their corresponding outputs. On the basis of Peng's work, many improved algorithms were proposed (Huang et al. 2014; Balasundaram and Meena 2016b; Khemchandani et al. 2016; Parastalooi et al. 2016; Rastogi et al. 2017). In this section, a novel projection wavelet weighted TSVR is proposed for deep fading channel estimation, which inherits the advantages of TSVR regression for nonlinear systems.

Given a training set $\mathbf{Tr} = \{(\mathbf{t}_1, z_1), (\mathbf{t}_2, z_2), \dots, (\mathbf{t}_m, z_m)\}$, where $\mathbf{t}_i \in \mathbb{R}^2$ and $z_i \in \mathbb{R}$, $i = 1, 2, \dots, m$, m is the number of training samples. Then the output vector of the training data can be denoted as $\mathbf{Z} = (z_1, z_2, \dots, z_m)^T \in \mathbb{R}^m$ and the input matrix as $\mathbf{T} = (\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_m)^T \in \mathbb{R}^{m \times 2}$. The first column of $\mathbf{T}$ represents the time domain positions of training samples and the second column of $\mathbf{T}$ represents the frequency domain positions of training samples. Let $\mathbf{e}$ and $\mathbf{I}$ be a ones vector and an identity matrix of appropriate dimensions, respectively.

In high mobility wireless environment, the channel is selective in both time and frequency domains, and the doubly selective channel exhibits very complex nonlinearity in the case of fast and deep fading. Therefore, linear channel estimation methods cannot obtain high performance. We adopt the nonlinear PWWTSVR algorithm to satisfy the estimation requirements of nonlinear channels since TSVR is superior in solving nonlinear, small training samples and high dimensional pattern recognition. Similar to the classical TSVR model (Peng 2010), the PWWTSVR is constructed by two nonparallel hyperplanes, down-bound $f_1(\mathbf{t})$, and up-bound $f_2(\mathbf{t})$; each hyperplane determines the ϵ -insensitive bound regressor, and the end regressor is $f(\mathbf{t}) = \frac{1}{2}(f_1(\mathbf{t}) + f_2(\mathbf{t}))$.

3.1 Channel estimation algorithm-projection wavelet weighted twin support vector regression

Initially, SVR was used to solve linear regression problems and a hyperplane can be adopted to regress the relationship between inputs and outputs by training samples. However, the hyperplane method is only applicable to linear problems. Based on Vapnik theory (1995), the algorithm can be extended to nonlinear cases by using kernel mapping, which is the majority of cases in the real world. The kernel trick is adopted to map the input into a higher-dimensional feature space, and the following kernel-generated functions are considered: down-bound $f_1(\mathbf{t}) = K(\mathbf{t}, \mathbf{T}^T)\mathbf{w}_1 + b_1$ and up-bound $f_2(\mathbf{t}) = K(\mathbf{t}, \mathbf{T}^T)\mathbf{w}_2 + b_2$, where $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^m$, m is the number of training points, $b_1, b_2 \in \mathbb{R}$, K is an appropriately chosen kernel. Therefore, the end regressor is the average of $f_1(\mathbf{t})$ and $f_2(\mathbf{t})$, i.e. $f(\mathbf{t}) = \frac{1}{2}(f_1(\mathbf{t}) + f_2(\mathbf{t}))$. The optimization problems can be described as follows:

$$\begin{aligned} \min_{\mathbf{w}_1, b_1, \xi_1} \quad & L_1 = \frac{1}{2}(\mathbf{Y} - (K(\mathbf{T}, \mathbf{T}^T)\mathbf{w}_1 + \mathbf{e}b_1))^T \mathbf{D}(\mathbf{Y} - (K(\mathbf{T}, \mathbf{T}^T)\mathbf{w}_1 + \mathbf{e}b_1)) + \frac{1}{2}c_{11}(\mathbf{w}_1^T \mathbf{w}_1 + b_1^2) \\ & + \frac{1}{2}c_{12}\hat{\mathbf{w}}_1^T \Sigma_z \hat{\mathbf{w}}_1 + c_{13}\mathbf{e}^T \mathbf{D} \xi_1 \\ \text{s.t.} \quad & \mathbf{Y} - (K(\mathbf{T}, \mathbf{T}^T)\mathbf{w}_1 + \mathbf{e}b_1) \geq -\varepsilon_1 \mathbf{e} - \xi_1, \quad \xi_1 \geq 0\mathbf{e}, \end{aligned} \quad (8)$$

and

$$\begin{aligned} \min_{\mathbf{w}_2, b_2, \xi_2} \quad & L_2 = \frac{1}{2}(\mathbf{Y} - (K(\mathbf{T}, \mathbf{T}^T)\mathbf{w}_2 + \mathbf{e}b_2))^T \mathbf{D}(\mathbf{Y} - (K(\mathbf{T}, \mathbf{T}^T)\mathbf{w}_2 + \mathbf{e}b_2)) + \frac{1}{2}c_{21}(\mathbf{w}_2^T \mathbf{w}_2 + b_2^2) \\ & + \frac{1}{2}c_{22}\hat{\mathbf{w}}_2^T \Sigma_z \hat{\mathbf{w}}_2 + c_{23}\mathbf{e}^T \mathbf{D} \xi_2 \\ \text{s.t.} \quad & (K(\mathbf{T}, \mathbf{T}^T)\mathbf{w}_2 + \mathbf{e}b_2) - \mathbf{Y} \geq -\varepsilon_2 \mathbf{e} - \xi_2, \quad \xi_2 \geq 0\mathbf{e}, \end{aligned} \quad (9)$$

where $c_{11}, c_{12}, c_{13}, c_{21}, c_{22}, c_{23} > 0$ are parameters chosen a priori by the user; ε_1 and ε_2 are insensitive parameters, ξ_1 and ξ_2 are slack vectors to measure the errors of samples outside the “ ε tube”. $\mathbf{D} \in \mathbb{R}^{m \times m}$ is a weighting matrix, which will be discussed later.

The first term in the objective function of (8) is the sum of weighted squared distances from training points to the down-bound function, which is called empirical risk. Minimizing this causes the function $f_1(\mathbf{t})$ to fit the training samples and avoid under-fitting. The second term is a regularization term, which can make $f_1(\mathbf{t})$ as smooth as possible. The structural risk minimization is implemented by minimizing the regularization term $\frac{1}{2}(\mathbf{w}_1^T \mathbf{w}_1 + b_1^2)$. A small value of $\frac{1}{2}(\mathbf{w}_1^T \mathbf{w}_1 + b_1^2)$ corresponds to the function (8) being flat. The third one, data structure term, can minimize empirical variance values of projected points on the down-bound functions. The fourth term aims to minimize the sum of errors of the points lower than the down-bound $f_1(\mathbf{t})$, which can possibly over-fit the training points. The ratios of the four penalty terms in the objective function of (8) can be adjusted by the choice of c_{11}, c_{12} , and c_{13} . The optimization problem (9) is with similar illustrations.

In the third term, the projection axis, $\hat{\mathbf{w}}_k = [\mathbf{w}_k; -1]$, $k = 1, 2$, is normal to the line of the bound regression functions. The projection axis is meant to make the projected zone or the variance of the projected noise as small as possible, and the following formula can be obtained.

$$\min \frac{1}{2}\hat{\mathbf{w}}_k^T \Sigma_z \hat{\mathbf{w}}_k = \min \frac{1}{2}\mathbf{w}_k^T \Sigma_{\varphi(\mathbf{t})} \mathbf{w}_k - \mathbf{w}_k^T \Sigma_{\varphi(\mathbf{t})} y \quad (10)$$

where

$$\begin{aligned} \Sigma_z &= \frac{1}{n} \sum_{i=1}^n (\mathbf{z}_i - \mu_z)(\mathbf{z}_i - \mu_z)^T, \\ \Sigma_{\varphi(\mathbf{t})} &= \frac{1}{n} \sum_{i=1}^n (\varphi(\mathbf{t}_i) - \mu_{\varphi(\mathbf{t})})(\varphi(\mathbf{t}_i) - \mu_{\varphi(\mathbf{t})})^T, \\ \Sigma_{\varphi(\mathbf{t})y} &= \frac{1}{n} \sum_{i=1}^n \varphi(\mathbf{t}_i)y_i - \mu_{\varphi(\mathbf{t})}\mu_y, \end{aligned} \quad (11)$$

and $\mathbf{z}_i$ is the training point $\mathbf{z}_i = (\varphi(\mathbf{t}_i); y_i)$, $i = 1, \dots, m$, μ_z is the centroid point of $\mathbf{z}_i$, and Σ_z is the covariance matrix of $\mathbf{z}_i$, $\Sigma_{\varphi(\mathbf{t})}$ is the empirical covariance matrices of inputs, $\Sigma_{\varphi(\mathbf{t})y}$ is the empirical correlation coefficient matrix between the inputs and responses. They are defined as $\mu_{\varphi(\mathbf{t})}$ and μ_y , which are the centroid points of the inputs and outputs respectively. $\varphi(\mathbf{t})$ is the operation of mapping input $\mathbf{t}$ to high-dimensional feature space using kernel trick for nonlinear applications, i.e., $\mathbf{t} \rightarrow \varphi(\mathbf{t})$. However, $\varphi(\mathbf{t})$ lacks an explicit formulation

due to the higher dimensions, which prevents the computation of $\Sigma_{\varphi(\mathbf{t})}$. In Peng's work (2014), the eigenvalue decomposition method is adopted to explicitly map to the empirical feature space. Let $K(\mathbf{T}, \mathbf{T}^T)$ denote an $m \times m$ matrix of rank r , where K is an appropriately chosen kernel. Since $K(\mathbf{T}, \mathbf{T}^T)$ is a symmetric positive-semidefinite matrix, it can be decomposed as $K(\mathbf{T}, \mathbf{T}^T) = \mathbf{P}_{m \times r} \mathbf{\Lambda} \mathbf{P}_{r \times m}^T$, where $\mathbf{\Lambda}$ is a diagonal matrix containing only the r positive eigenvalues of $K(\mathbf{T}, \mathbf{T}^T)$ in decreasing order, and $\mathbf{P}_{m \times r}$ consists of the eigenvectors corresponding to the positive eigenvalues. The mapping from the input data space to the kernel space is expressed as

$$\mathbf{t} \rightarrow \varphi(\mathbf{t}) = \mathbf{\Lambda}^{-1/2} \mathbf{P}^T K(\mathbf{t}, \mathbf{T}^T). \quad (12)$$

To solve the optimization problems (8) and (9), we convert the constrained problems to a pair of unconstrained problems by introducing the plus function $(\cdot)_+$ and substitute (10) into (8) and (9) as follows.

$$\begin{aligned} \min_{\mathbf{w}_1, b_1} L_1 = & \frac{1}{2} (\mathbf{Y} - (K(\mathbf{T}, \mathbf{T}^T) \mathbf{w}_1 + \mathbf{e} b_1))^T \mathbf{D} (\mathbf{Y} - (K(\mathbf{T}, \mathbf{T}^T) \mathbf{w}_1 + \mathbf{e} b_1)) + \frac{1}{2} c_{11} (\mathbf{w}_1^T \mathbf{w}_1 + b_1^2) \\ & + \frac{1}{2} c_{12} (\mathbf{w}_1^T \Sigma_{\varphi(\mathbf{t})} \mathbf{w}_1 - \mathbf{w}_1^T \Sigma_{\varphi(\mathbf{t})y}) + c_{13} e^T \mathbf{D} ((K(\mathbf{T}, \mathbf{T}^T) \mathbf{w}_1 + \mathbf{e} b_1) - (\mathbf{Y} + \varepsilon_1 \mathbf{e}))_+ \end{aligned} \quad (13)$$

and

$$\begin{aligned} \min_{\mathbf{w}_2, b_2} L_2 = & \frac{1}{2} (\mathbf{Y} - (K(\mathbf{T}, \mathbf{T}^T) \mathbf{w}_2 + \mathbf{e} b_2))^T \mathbf{D} (\mathbf{Y} - (K(\mathbf{T}, \mathbf{T}^T) \mathbf{w}_2 + \mathbf{e} b_2)) + \frac{1}{2} c_{21} (\mathbf{w}_2^T \mathbf{w}_2 + b_2^2) \\ & + \frac{1}{2} c_{22} (\mathbf{w}_2^T \Sigma_{\varphi(\mathbf{t})} \mathbf{w}_2 - \mathbf{w}_2^T \Sigma_{\varphi(\mathbf{t})y}) + c_{23} e^T \mathbf{D} ((\mathbf{Y} - \varepsilon_2 \mathbf{e}) - (K(\mathbf{T}, \mathbf{T}^T) \mathbf{w}_2 + \mathbf{e} b_2))_+ \end{aligned} \quad (14)$$

Plus functions $(\cdot)_+$ in (13) and (14) are not differentiable, but they can be replaced by smooth approximate functions $\mathbf{p}(\cdot)$. In this paper, we adopt the sigmoid integral function as a smooth function; it is defined as

$$\mathbf{p}(\mathbf{t}) = \mathbf{t} + \frac{1}{\alpha} \log(1 + \exp(-\alpha \mathbf{t})). \quad (15)$$

where α is a positive real constant. Define $\mathbf{f}_1 = (\mathbf{Y} + \varepsilon_1 \mathbf{e})$, $\mathbf{f}_2 = (\mathbf{Y} - \varepsilon_2 \mathbf{e})$, $\mathbf{u}_1 = [\mathbf{w}_1^T, b_1]^T$, $\mathbf{u}_2 = [\mathbf{w}_2^T, b_2]^T$, $\mathbf{G} = [K(\mathbf{T}, \mathbf{T}^T), \mathbf{e}]$ and replace the plus functions $(\cdot)_+$ in (13) and (14) by (15), we can get

$$\begin{aligned} \min_{\mathbf{w}_1, b_1} L_1 = & \frac{1}{2} (\mathbf{Y} - (K(\mathbf{T}, \mathbf{T}^T) \mathbf{w}_1 + \mathbf{e} b_1))^T \mathbf{D} (\mathbf{Y} - (K(\mathbf{T}, \mathbf{T}^T) \mathbf{w}_1 + \mathbf{e} b_1)) + \frac{1}{2} c_{11} (\mathbf{w}_1^T \mathbf{w}_1 + b_1^2) \\ & + \frac{1}{2} c_{12} (\mathbf{w}_1^T \Sigma_{\varphi(\mathbf{t})} \mathbf{w}_1 - \mathbf{w}_1^T \Sigma_{\varphi(\mathbf{t})y}) + c_{13} e^T \mathbf{D} \mathbf{p}(\mathbf{G} \mathbf{u}_1 - \mathbf{f}_1), \end{aligned} \quad (16)$$

and

$$\begin{aligned} \min_{\mathbf{w}_2, b_2} L_2 = & \frac{1}{2} (\mathbf{Y} - (K(\mathbf{T}, \mathbf{T}^T) \mathbf{w}_2 + \mathbf{e} b_2))^T \mathbf{D} (\mathbf{Y} - (K(\mathbf{T}, \mathbf{T}^T) \mathbf{w}_2 + \mathbf{e} b_2)) + \frac{1}{2} c_{21} (\mathbf{w}_2^T \mathbf{w}_2 + b_2^2) \\ & + \frac{1}{2} c_{22} (\mathbf{w}_2^T \Sigma_{\varphi(\mathbf{t})} \mathbf{w}_2 - \mathbf{w}_2^T \Sigma_{\varphi(\mathbf{t})y}) + c_{23} e^T \mathbf{D} \mathbf{p}(\mathbf{f}_2 - \mathbf{G} \mathbf{u}_2). \end{aligned} \quad (17)$$

Note that L_1 in (16) and L_2 in (17) are convex (Wang et al. 2019). Global and unique solutions can be obtained, and the Newton iterative approach can be adopted to solve the minimization problems as follows.

The first- and second-order gradients of L_1 in (16) and L_2 in (17) are deduced as follows

$$\nabla L_1(\mathbf{u}_1) = \mathbf{Q}_1 \mathbf{u}_1 - \mathbf{P}_1 + c_{13} \mathbf{G}^T \mathbf{D} \text{diag} \left(\frac{1}{1 + \exp(-\alpha(\mathbf{G} \mathbf{u}_1 - \mathbf{f}_1))} \right) \quad (18)$$

$$\nabla^2 L_1(\mathbf{u}_1) = \mathbf{Q}_1 + \alpha c_{13} \mathbf{G}^T \mathbf{D} \text{diag} \left(\frac{\exp(-\alpha(\mathbf{G} \mathbf{u}_1 - \mathbf{f}_1))}{1 + \exp(-\alpha(\mathbf{G} \mathbf{u}_1 - \mathbf{f}_1))^2} \right) \mathbf{G}, \quad (19)$$

and

$$\nabla L_2(\mathbf{u}_2) = \mathbf{Q}_2 \mathbf{u}_2 - \mathbf{P}_2 - c_{23} \mathbf{G}^T \mathbf{D} \text{diag} \left(\frac{1}{1 + \exp(-\alpha(\mathbf{f}_2 - \mathbf{G} \mathbf{u}_2))} \right) \quad (20)$$

$$\nabla^2 L_2(\mathbf{u}_2) = \mathbf{Q}_2 + \alpha c_{23} \mathbf{G}^T \mathbf{D} \text{diag} \left(\frac{\exp(-\alpha(\mathbf{f}_2 - \mathbf{G} \mathbf{u}_2))}{1 + \exp(-\alpha(\mathbf{f}_2 - \mathbf{G} \mathbf{u}_2))^2} \right) \mathbf{G}, \quad (21)$$

where

$$\begin{aligned} \mathbf{Q}_1 &= \begin{bmatrix} c_{11} \mathbf{I} + K(\mathbf{T}, \mathbf{T}^T)^T \mathbf{D} K(\mathbf{T}, \mathbf{T}^T) + \Sigma_{\varphi(\mathbf{t})} & K(\mathbf{T}, \mathbf{T}^T)^T \mathbf{D} \mathbf{e} \\ \mathbf{e}^T \mathbf{D} K(\mathbf{T}, \mathbf{T}^T) & c_{11} + \mathbf{e}^T \mathbf{D} \mathbf{e} \end{bmatrix}, \\ \mathbf{P}_1 &= \begin{bmatrix} K(\mathbf{T}, \mathbf{T}^T)^T \mathbf{D} \mathbf{Y} + c_{12} \Sigma_{\varphi(\mathbf{t})\mathbf{y}} \\ \mathbf{e}^T \mathbf{D} \mathbf{Y} \end{bmatrix}, \quad \mathbf{Q}_2 = \begin{bmatrix} c_{21} \mathbf{I} + \mathbf{A}^T \mathbf{D} K(\mathbf{T}, \mathbf{T}^T) + \Sigma_{\varphi(\mathbf{x})} & K(\mathbf{T}, \mathbf{T}^T)^T \mathbf{D} \mathbf{e} \\ \mathbf{e}^T \mathbf{D} K(\mathbf{T}, \mathbf{T}^T) & c_{21} + \mathbf{e}^T \mathbf{D} \mathbf{e} \end{bmatrix}, \\ \text{and } \mathbf{P}_2 &= \begin{bmatrix} K(\mathbf{T}, \mathbf{T}^T)^T \mathbf{D} \mathbf{Y} + c_{22} \Sigma_{\varphi(\mathbf{t})\mathbf{y}} \\ \mathbf{e}^T \mathbf{D} \mathbf{Y} \end{bmatrix}. \end{aligned}$$

The iterative solutions of minimization problems (13) and (14) can be obtained by adopting the Newton method and using (18)–(21), as follows:

$$\mathbf{u}_1^{k+1} = \mathbf{u}_1^k - (\nabla^2 L_1(\mathbf{u}_1))^{\text{--}1} (\nabla L_1(\mathbf{u}_1)), \quad (22)$$

$$\mathbf{u}_2^{k+1} = \mathbf{u}_2^k - (\nabla^2 L_2(\mathbf{u}_2))^{\text{--}1} (\nabla L_2(\mathbf{u}_2)). \quad (23)$$

3.2 Training samples

SVR is a supervised machine learning method, which requires input and output of training samples for parameter training. The inputs of training samples are the positions of estimated channel response in time and frequency domain, and the outputs are the channel response estimated by received pilots. In the proposed method, the pilots are inserted equidistantly in both time and frequency domain. The OFDM symbols inserted into the pilot are called pilot symbols, and those without pilot are called data symbols. The positions of pilot symbols are expressed as $[n\Delta t]$, $n = 0, 1, \dots, N_t - 1$, where Δt is the pilot symbol interval in time domain and N_t is the number of pilot symbols. In an OFDM symbol, the transmitting pilot subcarrier positions are expressed as $[m\Delta f]$, $m = 0, 1, \dots, N_f - 1$, where Δf is the pilot interval in frequency domain and N_f is the number of pilots in an OFDM symbol. The pilot insertion scheme is shown in Fig. 2. Let the transmitting pilot matrix be $\mathbf{X}_p = \text{diag}(X(n\Delta t, m\Delta f)) \in \mathbb{C}^{N_t N_f \times N_t N_f}$, and the channel frequency response estimated at pilot subcarriers according to (7) can be obtained

Fig. 2 Pilot insertion scheme in OFDM system

$$\hat{\mathbf{H}}_p = (\mathbf{X}_p)^{-1} \mathbf{Y}_p \quad (24)$$

where $\mathbf{Y}_p = \mathbf{Y}(n\Delta t, m\Delta f) \in \mathbb{C}^{N_t N_f}$ is the received pilot vector and $\hat{\mathbf{H}}_p = \hat{\mathbf{H}}(n\Delta t, m\Delta f) \in \mathbb{C}^{N_t N_f}$ is the estimated frequency response at pilot positions $(n\Delta t, m\Delta f)$.

After interpolation, the frequency response of data position can be calculated and the predicted frequency response of all subcarriers in an OFDM symbol can be expressed as

$$\tilde{\mathbf{H}} = f(\hat{\mathbf{H}}_p). \quad (25)$$

3.3 Weighting parameters

The parameter $\mathbf{D}$ mentioned above is the weighting matrix, and it can be determined beforehand according to the importance of training data. The samples used for training are all polluted by noise of different amplitude. Intuitively, the samples polluted by large noise should be given smaller weight. Therefore, all training data should be given penalty weights. The penalty weighting parameter D are given to the samples in the proposed algorithm. The weighting parameters which represent the distance between noised samples and its 'real position' can reflect the prior information of the training samples. While in the classical SVR algorithms, such as the TSVR, no weights are added to the samples, that means the weights are all one and all of the samples have the same weights. Points with too much noise, such as outliers, will degrade the performance of the regressors. Different training samples should be given different weights, a larger given weight means that the sample is more important. Motivated by this idea and noting that the Gaussian function

can reflect this trend, the weighting parameter $\mathbf{D}$ is determined by the Gaussian function described as follows.

$$\mathbf{D} = \text{diag}\left(Ae^{(-|\mathbf{Y}-\hat{\mathbf{Y}}|^2/\sigma^2)}\right), \quad (26)$$

where A is the amplitude, σ represents the standard deviation, and $\hat{\mathbf{Y}} (= [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{N_t N_f}]^T)$ is the estimation value vector of output $\mathbf{Y} (= [y_1, y_2, \dots, y_{N_t N_f}]^T)$. In this paper, wavelet transform based method is adopted to calculate $\hat{\mathbf{Y}}$. Wavelet transform is a kind of time-frequency transform analysis method and wavelet filter is applicable to short-term signal processing, which also is the characteristic of most actual time series signals. It inherits and develops the idea of short-time Fourier transform localization, and overcomes the shortcomings of window size not changing with frequency. It can provide a time-frequency window that changes with frequency. It is a relatively ideal tool for time-frequency analysis and processing signals. In addition, the inserted pilots are two-dimensional distribution in time and frequency domain, therefore two-dimensional wavelet filtering is adopted to calculate $\hat{\mathbf{Y}}$ by three steps. Firstly, the wavelet transform may be considered a form of time-frequency representation for signals, and thus are related to harmonic analysis. For discrete data, discrete wavelet transforms (DWT) use discrete-time filter banks to decompose signal sequence. Secondly, the obtained decomposed sequence is processed by an appropriate algorithm to remove noise. In this paper, the high frequency part of the frequency domain signal is directly zeroed as denoising algorithm. Thirdly, the estimation value of output $\hat{y}$ is reconstructed by the denoised sequence and the estimation value of output $\hat{\mathbf{Y}}$ can be obtained. Substitute $\hat{\mathbf{Y}}$ into (26), and then the weighting vector $\mathbf{d}$ and matrix $\mathbf{D}$ can be calculated.

Weighting parameters were also introduced in other improved TSVR, for example, in Xu's work (2014), the K-nearest neighbor (KNN) algorithm is adopted. One of the differences between the PWWTSSVR and the method proposed by Xu and Wang (2014) lies in the different objects being processed. The KNN algorithm is suitable for dealing with clustered samples. The idea of the KNN is that a sample point x is important if it has a larger number of K-nearest neighbors, whereas it is not important if it is an outlier that has a small number of K-nearest neighbors. While the wavelet transform theory adopted in our proposed algorithm is a kind of time-frequency representation for signals, and the proposed method based on the wavelet theory is suitable for dealing with time series samples.

3.4 Algorithm summary

In this subsection, we illustrate the channel estimation algorithm.

Algorithm: Channel estimation based on PWWTSSVR

Input:

- δ : the error tolerance for learning accuracy.
- M : the maximum number of iterations.
- $\mathbf{X}_P$: transmitted pilot matrix, $\mathbf{Y}_P$ received pilot vector, $\mathbf{T} = [n\Delta t, m\Delta f]$ pilot position vector
- $c_{11}, c_{12}, c_{13}, c_{21}, c_{22}, c_{23}$: parameters in (16) and (17), A, σ^2 in (26), α in (15)
- $\mathbf{u}_1^0 = \mathbf{e}, \mathbf{u}_2^0 = \mathbf{e}$: initial approximation of $\mathbf{u}_1$ and $\mathbf{u}_2$.
- $\mathbf{f}_1 = (\mathbf{Y} + \varepsilon_1 \mathbf{e}), \mathbf{f}_2 = (\mathbf{Y} - \varepsilon_2 \mathbf{e})$.

Output: The optimal solution $\mathbf{u}_1, \mathbf{u}_2$.

Process:

1. Calculate channel frequency response at pilot subcarriers $\hat{\mathbf{H}}_P$ by (24).
2. $\mathbf{Y} = \text{real}(\hat{\mathbf{H}}_P)$, where $\text{real}(\cdot)$ means taking real part. $\mathbf{Y}$ is denoised by wavelet transform, and then $\hat{\mathbf{Y}}$ is obtained. Calculate $\mathbf{d}$ by (26).
3. Calculate $\Sigma_{\varphi(\mathbf{t})}$ and $\Sigma_{\varphi(\mathbf{t})y}$ by (11).
4. $k = 0$.

do{

Calculate $\nabla L_1(\mathbf{u}_1), \nabla^2 L_1(\mathbf{u}_1), \nabla L_2(\mathbf{u}_2)$ and $\nabla^2 L_2(\mathbf{u}_2)$ by (18)-(21)

· Solve $\mathbf{u}_1^{k+1}$ and $\mathbf{u}_2^{k+1}$ by (22) and (23)

· $err1 = \text{sum}(\mathbf{u}_1^{k+1} - \mathbf{u}_1^k), err2 = \text{sum}(\mathbf{u}_2^{k+1} - \mathbf{u}_2^k)$, where $\text{sum}(\cdot)$ means taking the sum of vector elements.

· $k = k + 1$

while $((|err1| > \delta \text{ or } |err2| > \delta) \text{ and } k < M)$

5. Compute $h_{\text{real}}(\mathbf{t}) = \frac{1}{2} K(\mathbf{t}, \mathbf{T}^T)(\mathbf{w}_1 + \mathbf{w}_2)^T + \frac{1}{2}(b_1 + b_2)$.

6. Let $\mathbf{Y} = \text{imag}(\hat{\mathbf{H}}_P)$, where $\text{imag}(\cdot)$ means taking imaginary part, repeat 2-5 once, $h_{\text{imag}}(\mathbf{t})$ can be obtained.

7. Frequency response of all subcarriers in OFDM symbols can be expressed as $(\tilde{\mathbf{H}})_{\mathbf{t}} = h_{\text{real}}(\mathbf{t}) + h_{\text{imag}}(\mathbf{t})j$, where $(\tilde{\mathbf{H}})_{\mathbf{t}}$ means the $\mathbf{t} = (n, m)$ element of $\tilde{\mathbf{H}}$.

3.5 Computational complexity analysis

Computational complexity is an important performance of an algorithm. Consider that in a calculating period, if the number of pilot symbols is N_t and the number of pilots in one pilot symbol is N_f , then the number of training samples is $l = N_t N_f$. In the processing step 1 of the algorithm described in the previous subsection, the calculation of $\hat{\mathbf{H}}_p$ requires $O(l)$ complex multiplications, and the step 2 needs computational cost $O(l)$ real multiplication. In step 3, since the input $\mathbf{t}$ is known at receiver, the $\varphi(\mathbf{t})$ in (12) can be pre-computed and stored at the receiver. Therefore the calculation of $\Sigma_{\varphi(\mathbf{t})}$ and $\Sigma_{\varphi(\mathbf{t})\mathbf{y}}$ require $O(l)$ and $O(l)$ respectively. The step 4 is an iterative process. The main calculation comes from the inversion of two matrices: $(\nabla^2 L_1(\mathbf{u}_1))^{-1}$ and $(\nabla^2 L_2(\mathbf{u}_2))^{-1}$ in (22) and (23) respectively, and one matrix inversion requires about $O(l^3)$. Let the number of iteration steps be n (generally, $n < 10$), then the computational complexity of this step is about $O(2nl^3)$. In addition, the proposed algorithm contains the wavelet transform weighted matrix, and a Db-3 wavelet with a length of 6 is used in this paper, then the complexity of the wavelet transform is less than $12l$. By comparing with the computations of the inverse matrix, the complexity of computing the wavelet matrix can be ignored. The step 5 also needs computational cost $O(l)$. The step 6 is the repetitive process of the previous calculation. If we only retain the inverse computations and ignore the other computations, the total computational complexity is about $O(4nl^3)$.

4 Experimental results

In this section, we show some simulation results of the proposed channel estimation algorithm based on the novel PWWT-SVR. TSVR is effective for channel estimation, and its regression performance has been verified by the works of Charrada and Samet (2016) and Matilde et al. (2004), therefore we compare performance of the proposed algorithm with the linear interpolation method and TSVR (Peng 2010) based channel estimation algorithm. Consider an OFDM system with doubly selective channel. The multipath number $L + 1 = 6$, and the channel taps are assumed to be i.i.d., correlated in time with a correlation function according to Jakes' model (Stuber 1996) $E[h(n_1, l_1)h^*(n_2, l_2)] = \sigma_h^2 J_0(2\pi f_{\max} T_s (n_1 - n_2)) \delta(l_1 - l_2)$, where $E(\cdot)$ means expected value, $(\cdot)^*$ denotes conjugate, n_k is time index, l_k is channel path index, J_0 is the zeroth-order Bessel function of the first kind, T_s is the sampling time interval, and σ_h^2 denotes the variance of the channel. We consider $N = 64$ subcarriers, and a CP of length $L = 5$. The sampling time interval $T_s = 72\mu s$, carrier frequency $f_c = 2.15$ GHz, mobile speed is 120/350 (km/h) and 16-QAM signaling is assumed.

In order to demonstrate the effects of multipath and moving speed on channel frequency response, four scenario simulations are performed. Figure shows the channel frequency response at subcarriers in an OFDM symbol under multipath number being 2 and 5 for mobile speed equal to 0 and 350 km/h respectively. From Fig. 3, we can see that multipath can cause frequency-domain fading and the more the number of multipath, the deeper the fading. At the same time, the mobile can cause ICI, the faster the movement, the bigger the ICI. According to (7), we know that the ICI is influenced by the product of channel response $\mathbf{H}_n$ and data to be transmitted $\mathbf{X}$. As data to be transmitted is random, the ICI caused by transmitted data is like noise, which can be reflected in the simulations.

Fig. 3 Channel frequency response at subcarriers in an OFDM symbol under multipath number being 2

The computer simulations are implemented in a Matlab R2014a environment on a PC. Because Gaussian kernel is an effective and frequently used kernel function in TSVR research field, in this paper, the Gaussian nonlinear kernel is adopted for the proposed PWWTSVR and TSVR (Peng 2010), that is

$$K(\mathbf{x}^T, \mathbf{y}^T) = \exp \left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{\rho} \right), \quad (27)$$

where ρ determines the width of the Gaussian function. The choice of parameters is essential for the performance of algorithms. In this paper, parameter values are chosen by the grid search method from the set of values $\{10^i | i = -4, -3, \dots, 5\}$.

Figure 4 illustrates the regression of PWWTSVR. In this simulation experiment, pilots are inserted in frequency and time domain with insertion interval $\Delta f = 2$, $\Delta t = 2$, SNR=10dB, number of multipath $L = 5$ and mobile speed $v = 350\text{km/h}$. We can see that the proposed PWWTSVR can fit the channel frequency response and the outliers (pilot samples with too much noise) can be ignored, which shows the robustness of the proposed algorithm. The star points represent the estimated channel response at the pilot position by LS method, the solid line represents the interpolation using the PWWTSVR algorithm, and the dashed line represents the perfect estimation result, which is not affected by noise in the simulation.

The performance of channel estimation based on PWWTSVR and other methods are evaluated by selected criteria. The number of testing samples is denoted by l , y_i denotes the real value of a testing sample point x_i and $\hat{y}_i$ denotes the predicted value of x_i . The criteria are specified as follows.

SNR: Signal-to-noise ratio, defined as $\text{SNR} = 10 \log(\sigma_x^2 / \sigma_v^2)$, where the signal power $\sigma_x^2 = E(|x(k)|^2)$, σ_v^2 is variance of AWGN.

SSE: Sum squared error of testing samples, defined as $\text{SSE} = \sum_{i=1}^l |y_i - \hat{y}_i|^2$.

BER: bit error rate, defined as $\text{BER} = N_e / N$, where N_e and N are the number of error signals and all signals in binary respectively.

Fig. 4 Performance of PWWTSSVR estimation method

SSE represents the fitting precision, but too small of an SSE value may mean overfitting of the regressor due to the fitting of noise.

To test the performance of kernel functions, six kernel functions are studied for Sinc function ($f(x) = \frac{\sin x}{x}$) including the Gaussian kernel adopted in the proposed method. The kernel function expressions and SSE performance are listed in Table 1. It is easy to see that the Gaussian kernel achieves the best results.

Outliers may have a certain influence on the regression of the method. To test the robustness of the proposed algorithm to outliers, the performance on Sinc function polluted by Gaussian additive noise with mean zero and variance $\zeta^2 = 0.1^2$ is studied. Figure 5 shows that the prediction performance of the proposed method is satisfying.

Figures 6, 7, 8 and 9 show the BER performance as a function of the SNR with different pilot spacing and different moving speed. We compare the BER variation of the proposed algorithm with TSVR estimation (Peng 2014), linear interpolation estimation, K-nearest TSVR (Gupta 2017) and perfect estimation in the presence of additive Gaussian noise. Table 2 shows the SSE performance of channel frequency response for various algorithms. It can be seen that the proposed method outperforms the TSVR estimation and linear interpolation estimation. The BER performances of the proposed method and the TSVR estimation are all better than that of linear interpolation estimation, which means that SVR

Table 1 Comparisons of PWWTSSVR on Sinc function with different kernel functions

Kernel Name	Function definition	SSE
Gaussian	$\exp(-\ x-x_i\ ^2/e)$	0.1458
Linear	xx_i^T	21.171
Polynomial	$(axx_i^T + c)^d$	13.602
Sigmoid	$\tanh(axx_i^T + c)$	0.4815
Multiquadric	$(\ x-x_i\ ^2 + \sigma^2)^{0.5}$	0.2754
Log	$-\log(1 + \ x-x_i\ ^\sigma)$	0.2146

Fig. 5 Regression performance of PWWTSSVR on noised Sinc function with outliers

Fig. 6 BER versus SNR for mobile speed $v = 120\text{km/h}$, pilot spacing $\Delta f = 2$, $\Delta t = 2$

is suitable for the regression problems based on training samples especially in the nonlinear cases. At the same time, we can see that our PWWTSSVR estimation outperforms the TSVR estimation, which shows the effectiveness of the adoption of the weighting parameter $\mathbf{D}$ in the objective functions.

Mobile speed v , pilot spacing Δf in frequency domain and Δt in time domain also affect the estimation performance. Slow moving speed leads to better BER and SSE performance, and smaller spacing interval can achieve better estimation quality, but the cost is the increase of computational complexity. The simulation results listed in table shows that our proposed method obtains the best performance compared with

Fig. 7 BER versus SNR for mobile speed $v = 120\text{km/h}$, pilot spacing $\Delta f = 4$, $\Delta t = 4$

Fig. 8 BER versus SNR for mobile speed $v = 350\text{km/h}$, pilot spacing $\Delta f = 2$, $\Delta t = 2$

TSVR estimation and linear interpolation estimation in the sense of SSE value, which shows its good fitting ability. Furthermore, we notice that there are special cases due to randomness. For example, the SSE of $\Delta f = 4$, $\Delta t = 4$, SNR=5 dB, $v = 350$ in Table 1 is less than that of $v = 120$.

In summary, the simulation results confirm that the proposed channel estimation algorithm has better performance than the conventional methods and the introduction of the projection item and the weighting parameters calculated by wavelet transform method can improve the performance of the channel estimation.

Fig. 9 BER versus SNR for mobile speed $v = 350\text{km/h}$, pilot spacing $\Delta f = 4$, $\Delta t = 4$

Table 2 SSE values of channel response by three estimation methods

SNR	0	5	10	15	20	25	30
$v = 120\text{ km/h}$, $\Delta t = 2$, $\Delta f = 2$							
Linear interpolation	3914.8	1152.7	395.68	115.01	43.397	20.903	15.949
TSVR	3072.3	1040.4	332.67	111.12	36.191	13.544	6.2251
K-nearest	2535.4	767.38	289.51	99.826	33.462	12.511	5.8285
PWWTSSR	1026.3	344.06	201.81	92.423	31.404	11.780	5.6006
$v = 350\text{km/h}$, $\Delta t = 2$, $\Delta f = 2$							
Linear interpolation	3944.4	1158.2	396.43	121.47	47.511	21.482	16.899
TSVR	3162.5	1066.0	338.27	112.53	41.681	15.686	6.8044
K-nearest	2853.2	898.15	302.35	100.07	37.283	14.702	6.6086
PWWTSSR	1070.0	353.17	202.21	95.074	35.714	13.346	6.1867
$v = 120\text{km/h}$, $\Delta t = 4$, $\Delta f = 4$							
Linear	3975.5	1583.5	553.36	314.36	187.00	160.92	133.67
TSVR	3910.0	1387.7	480.35	270.39	94.084	34.871	25.566
K-nearest	3532.3	1281.8	463.55	267.08	93.922	34.125	25.202
PWWTSSR	2841.3	1071.6	435.40	261.32	93.828	33.277	24.254
$v = 350\text{km/h}$, $\Delta t = 4$, $\Delta f = 4$							
Linear interpolation	4459.8	1607.2	554.70	340.49	219.81	173.54	101.80
TSVR	4197.2	1306.9	498.71	287.42	95.475	40.442	26.933
K-nearest	3895.5	1290.1	480.06	273.35	94.014	40.311	26.185
PWWTSSR	3287.4	1114.4	450.95	275.10	94.207	40.006	25.285

5 Conclusions

In this paper, a novel PWWTSSVR based channel estimator is proposed. Pilots are grid insertion in time and frequency domains in OFDM system, so the channel estimation is performed in dual domains using the grid-inserted pilots. Firstly, the channel frequency response is calculated at the pilot positions and the training samples for the PWWTSSVR can be obtained. Then channel frequency response at data positions can be estimated by the proposed algorithm. Unlike TSSVR, samples in different positions in the proposed PWWTSSVR model are given different weights according to the distance between samples and results preprocessed by the wavelet transform. Computational comparisons between PWWTSSVR and other existing methods are performed to show the better performance of the PWWTSSVR based channel estimation, which also demonstrates the effectiveness of the proposed method especially for the nonlinear cases. Furthermore, since one of the theoretical bases of the proposed method is wavelet theory, which is a powerful denoising tool for time series signals, the proposed method is suitable for dealing with time series datasets such as the estimation of channel parameters. Additionally, SVR is suitable for small sample size datasets, and a large number of training samples will bring about tremendous computational cost. For future work, study of small proportion pilots insertion in time and frequency domain algorithms based on more accurate channel models should be carried out.

Acknowledgements This work was supported in part by the Fund of the Scientific Research Project of Liaoning Provincial Department of Education (Grant No. 2019LNCJ02), in part by the Funds of the National Natural Science Foundation of China (Grant Nos. 71571091 and 71771112), and in part by the Fund of the University of Science and Technology Liaoning Talent Project (Grant No. 601011507-03).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anagha P, Balasundaram S, Meena Y (2018) On robust twin support vector regression in primal using squared pinball loss. *J Intell Fuzzy Syst* 35(5):5231–5239
- Anand P, Rastogi R, Chandra S (2018) Generalized ϵ -loss function-based regression. In: *Machine intelligence and signal analysis*, pp 395–409
- Anand P, Rastogi R, Chandra S (2019) A new asymmetric ϵ -insensitive pinball loss function based support vector quantile regression model. [arXiv:1908.06923](https://arxiv.org/abs/1908.06923)
- Balasundaram S, Meena Y (2016a) A new approach for training Lagrangian support vector regression. *Knowl Inf Syst* 49:1097–1129
- Balasundaram S, Meena Y (2016b) Training proximal twin support vector regression via unconstrained convex minimization. *Appl Intell* 44:931–955
- Balasundaram S, Gupta D, Kapil (2014) Lagrangian support vector regression via unconstrained convex minimization. *Neural Netw* 5:67–79
- Charrada A, Samet A (2016) Joint interpolation for LTE downlink channel estimation in very high-mobility environments with support vector machine regression. *IET Commun* 10(17):2435–2444
- Djouama A, Lim MS, Ettoumi FY (2014) Channel estimation in long term evolution uplink using minimum mean square error-support vector regression. *Wirel Pers Commun* 79:2291–2304

- Gao C, Shen M, Liu X, Wang L, Chu M (2019) End-point static control of basic oxygen furnace (BOF) steelmaking based on wavelet transform weighted twin support vector regression. *Complexity* 2019:1–16
- Gupta D (2017) Training primal K-nearest neighbor based weighted twin support vector regression via unconstrained convex minimization. *Appl Intell* 47(3):962–991
- Gupta U, Gupta D (2019) An improved regularization based Lagrangian asymmetric ν -twin support vector regression using pinball loss function. *Appl Intell* 49:3606–3627
- Huang XL, Shi L, Pelckmans K, Suykens JAK (2014) Asymmetric ν -tube support vector regression. *Comput Stat Data Anal* 77:371–382
- Khemchandani R, Goyal K, Chandra S (2016) TWSVR: regression via twin support vector machine. *Neural Netw* 74:14–21
- Matilde SF, Mario PC, Jeronimo AG, Fernando PC (2004) SVM multiregression for nonlinear channel estimation in multiple-input multiple-output systems. *IEEE Trans Signal Proc* 52(8):2298–2307
- Melki G, Cano A, Kecman V, Ventura S (2017) Multi-Target support vector regression via correlation regressor chains. *Inform Sci* s415–416:53–69
- Parastalooi N, Amiri A, Aliherdari P (2016) Modified twin support vector regression. *Neurocomputing* 211:84–97
- Peng X (2010) TSVR: an efficient twin support vector machine for regression. *Neural Netw* 23(3):356–372
- Peng X, Xu D, Shen J (2014) A twin projection support vector machine for data regression. *Neurocomputing* 138:131–141
- Rastogi R, Anand P, Chandra S (2017) A ν -twin support vector machine based regression with automatic accuracy control. *Appl Intell* 46:670–683
- Sheng Z, Tuan HD, Nguyen HH, Fang Y (2017) Pilot optimization for estimation of high-mobility OFDM Channels. *IEEE Trans Veh Technol* 66(10):8795–8806
- Stuber GL (1996) Principles of mobile communication. Kluwer, Boston
- Vapnik VN (1995) The natural of statistical learning theory. Springer, New York
- Wang LD, Gao C, Zhao NN, Chen XB (2019) A projection wavelet weighted twin support vector regression and its primal solution. *Appl Intell*. <https://doi.org/10.1007/s10489-019-01422-7>
- Wu J, Fan P (2016) A survey on high mobility wireless communications: Challenges, opportunities and solutions. *IEEE Access* 4:450–476
- Xu Y, Wang L (2014) K-nearest neighbor-based weighted twin support vector regression. *Appl Intell* 41:299–309
- Xu Y, Li X, Pan X, Yang Z (2017) Asymmetric ν -twin support vector regression. *Neural Comput Appl* 2:1–16
- Yu X, Luo Y, Chen X (2016) An optimized seamless dual-link handover scheme for high-speed rail. *IEEE Trans Veh Technol* 65(10):8658–8668

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Lidong Wang¹ · Yimei Ma¹ · Xudong Chang¹ · Chuang Gao¹ · Qiang Qu¹ · Xuebo Chen¹

Yimei Ma
1602776507@qq.com

Xudong Chang
1453923186@qq.com

Chuang Gao
13500422153@163.com

Qiang Qu
askdq@163.com

Xuebo Chen
xuebochen@126.com

- ¹ School of Electronic and Information Engineering, University of Science and Technology
Liaoning, Anshan, Liaoning, People's Republic of China