Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 11(1): 462–482.
DOI: 10.3934/math.2026020
Received: 05 November 2025
Revised: 21 December 2025
Accepted: 21 December 2025
Published: 06 January 2026

Research article

Support vector regression with orthogonal constraints for outlier suppression

Felix Ndudim[1] **and Thanasak Mouktonglang**[1,2,3,4,*]

[1] Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

[2] Data Science Research Center, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

[3] Advanced Research Center for Computational Simulation, Chiang Mai University, Chiang Mai 50200, Thailand

[4] Centre of Excellence in Mathematics, MHESI, Bangkok 10400, Thailand

* **Correspondence:** Email: thanasak.m@cmu.ac.th.

Abstract: This paper introduces a novel support vector regression framework with an orthogonal constraint (OC-SVR) for outlier suppression. The proposed method detects outliers via angular deviations in the feature space and modifies the support vector regression (SVR) optimization problem by enforcing orthogonality between the regression function and the detected outliers' feature directions, thereby neutralizing their influence. Experimental evaluations on benchmark datasets demonstrate that OC-SVR achieves significantly improved robustness and predictive accuracy compared to standard SVR and robust SVR variants.

Keywords: robust regression, support vector regression, outlier detection, orthogonal constraint, machine learning
Mathematics Subject Classification: 65K05, 90C20

1. Introduction

Support vector regression (SVR), designed by Vapnik [19], is one of the major tools for handling regression tasks due to its ability to model non-linearity and its practical performance even in high-dimensional space. Despite its effectiveness, SVR faces challenges and underperformance due to outliers and noisy data,since its optimization problem is based on the ε-insensitive loss [15, 21], which obviously does not differentiate between inliers and noisy data points during model training.

Several robust variants of SVR have been proposed in the literature. Some of the approaches modify the loss function (for example, Huber loss [7, 9], truncated loss [3]), while others reweight training samples to mitigate outlier influence [11, 13]. For example, a recent work suggests a robust weighted SVR using Hampel's weight function to improve predictive modeling under noisy conditions [11]. Local versions of support vector machine and support vector regression (SVM/SVR) with iterative weighting by residuals have also been introduced to reduce the impact of outliers in engineering forecasting applications [8, 10].

The latest developments in the literature include HawkEye loss, which designs an SVR variant with a bounded, smooth, and insensitive loss function, which helps to reduce large errors in a stable manner [1]. Another approach is the granular ball SVR (GBSVR), which groups data into balls to ease data structure and reduce outlier sensitivity. Support vector regression with sample significance (SSVR) assigns varying penalty weights to samples based on significance, thus down-weighting likely outliers [16, 18]. Additionally, efficient optimization methods have been applied in robust SVR , such as the alternate direction method of multipliers (ADMM) combined with Huber loss, to scale better in the presence of an outlier [5, 14, 20].

However, most of these methods still have drawbacks. Many rely on heuristic weighting schemes or require prior knowledge of noise distribution or thresholds, which can limit their generalizability. Loss function modifications (while powerful) may be computationally intensive to run. Approaches that detect outliers separately (outside the main optimization) do not always guarantee that their influence is suppressed in the learned model.

In this paper, we propose a novel method, support vector regression with orthogonal constraint for outlier suppression (OC-SVR). Our approach detects outliers in the reproducing kernel hilbert space (RKHS) using angular deviations from the centroid of the feature-mapped training points. We then modify the SVR optimization by introducing an orthogonal constraint that forces the regression function to be orthogonal to the feature directions of those detected outliers, effectively nullifying their influence. Experiments demonstrate that OC-SVR achieves superior robustness compared to standard SVR and existing robust SVR variants.

The main contributions of this paper are:

- A novel outlier detection criterion based on angular deviation from the centroid of the feature-mapped training points, which identifies anomalous samples in the RKHS more effectively than simple distance measures.
- A new robustness mechanism for SVR that enforces collective orthogonality between the regression function and the feature directions of detected outliers. This constraint cancels the net influence of outliers in feature space, offering a principled alternative to existing robust SVR methods that rely on reweighting, modified loss functions, or data removal.
- Demonstration of enhanced robustness and predictive accuracy relative to standard SVR and state-of-the-art robust SVR methods on benchmark datasets.

Glossary of terms

SVR (support vector regression): A kernel-based regression method derived from statistical learning theory, designed to model nonlinear relationships by mapping input data into a high-dimensional feature space.

OC-SVR (orthogonal constraint support vector regression): The proposed variant of support vector regression that incorporates an orthogonality constraint to suppress the influence of detected outliers in the feature space.

RKHS (reproducing kernel hilbert space): A Hilbert space induced by a positive definite kernel, in which inner products can be computed implicitly through kernel functions, enabling nonlinear learning.

Angular deviation: A geometric measure that quantifies the directional separation between a data point and the centroid in the reproducing kernel hilbert space (RKHS).

Outlier: A data point exhibiting an unusually large angular deviation from the data centroid, indicating anomalous behavior in the feature space.

RBF kernel (radial basis function kernel): A commonly used kernel defined as

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right),$$

with $\gamma > 0$ controlling the kernel bandwidth.

MSE (mean squared error): A performance metric defined as the average of the squared differences between predicted and true values.

RMSE (root mean squared error): The square root of the mean squared error, providing an error measure in the same scale as the target variable.

CCPP (combined cycle power plant): A Combined Cycle Power Plant dataset is a structured collection of operational measurements recorded from a combined cycle power plant, typically used for data analysis, regression, prediction, and machine-learning research. The dataset captures how operating conditions affect electrical power output and overall plant performance.

MAD (median absolute deviation): Median Absolute Deviation is a robust statistical measure of variability that quantifies how spread out data are around the median, rather than the mean. Because it relies on medians, MAD is highly resistant to outliers and non-Gaussian noise.

QP (quadratic program): Quadratic Programming is a class of mathematical optimization problems in which the objective function is quadratic and the constraints are linear. QP plays a central role in optimization theory and is widely used in machine learning, control, finance, and engineering.

KKT (Karush–Kuhn–Tucker): The Karush–Kuhn–Tucker conditions are a set of first-order optimality conditions that characterize optimal solutions of constrained optimization problems. They generalize the method of Lagrange multipliers to problems with inequality constraints and form the theoretical foundation of convex optimization, quadratic programming, and machine learning algorithms such as SVM and SVR.

PSD (positive semidefinite): Positive semidefinite is a property of a symmetric matrix that guarantees the associated quadratic form is non-negative. This concept is fundamental in optimization, linear algebra, statistics, and machine learning, especially in convex analysis.

2. Methodology

2.1. Outlier identification

The outlier detection strategy in OC-SVR is based on a geometric interpretation within the reproducing kernel hilbert space (RKHS). The first step is to compute the centroid of the mapped training data:

$$c = \frac{1}{n}\sum_{i=1}^{n}\phi(x_i), \tag{2.1}$$

where $\phi(x_i)$ denotes the feature mapping of the training point x_i into the RKHS, and n is the number of training samples. This centroid acts as a central reference, representing the average position of the data in the transformed feature space.

Since explicit computation of $\phi(x_i)$ is generally infeasible due to its high (or infinite) dimensionality, the kernel trick is utilized. Specifically, inner products are expressed as

$$\phi(x_i)\cdot\phi(x_j) = K(x_i, x_j), \tag{2.2}$$

where $K(\cdot,\cdot)$ is the kernel function (here, the RBF kernel). Consequently, the similarity between a mapped point $\phi(x_i)$ and the centroid is given by

$$\phi(x_i)\cdot c = \frac{1}{n}\sum_{j=1}^{n}K(x_i, x_j), \tag{2.3}$$

which leverages the kernel matrix constructed from all pairwise interactions.

The angular deviation of each point from the centroid is then computed as

$$\theta_i = \cos^{-1}\left(\frac{\phi(x_i)\cdot c}{||\phi(x_i)||||c||}\right), \tag{2.4}$$

where $||\phi(x_i)|| = \sqrt{K(x_i, x_i)}$ denotes the norm of the mapped point, and the centroid norm is obtained from

$$||c||^2 = \frac{1}{n^2}\sum_{i,j}K(x_i, x_j). \tag{2.5}$$

This angular measure quantifies the directional separation between each data point and the centroid. Points with higher angular deviations (θ_i) are considered potential outliers, as they exhibit significant directional separation from the data centroid in the RKHS.

2.2. Outlier threshold selection

To detect outliers, we used an angular deviation measure that quantifies the directional separation between each data point and the centroid in feature space. Instead of fixing the threshold a priori, we evaluated different percentile cut-offs of the angular deviation distribution on the training set. For each candidate threshold, data points with angular deviation values θ_i exceeding the cut-off were classified as outliers, forming the set O. To identify the most effective threshold, we trained the model under varying percentile thresholds (ranging from the 95th to the 80th) and computed the MSE for each case. Figure 1 illustrates the sensitivity of training performance to the angular-deviation threshold, highlighting that overly aggressive outlier removal degrades prediction accuracy. As shown in Figure 1, the MSE varied considerably across thresholds.

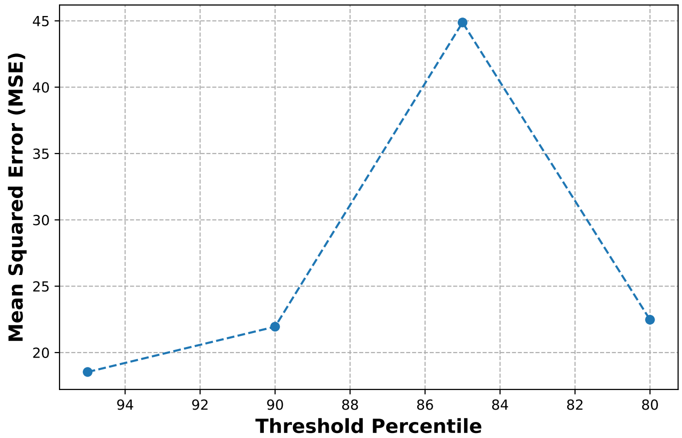

Figure 1. The MSE of OC-SVR on the training set of the CCPP dataset as a function of the angular-deviation threshold percentile. Each point corresponds to a model trained by treating samples with angular deviation above the given percentile as outliers and enforcing the orthogonal constraint. The minimum MSE is achieved at the 95th percentile, motivating the chosen threshold.

The 95th percentile threshold yielded the lowest error (MSE $\approx$ 18.6), while intermediate thresholds such as the 85th percentile resulted in substantially higher errors (MSE $\approx$ 45).

This analysis demonstrates that threshold selection has a direct impact on predictive performance. Based on empirical validation, we adopted the 95th percentile

$$\theta_{\text{th}} = \text{Percentile}_{95}\Big(\{\theta_i\}_{i=1}^{n}\Big), \tag{2.6}$$

as the optimal threshold, since it minimized the MSE in the training set. This approach provides a data-driven and performance-based criterion for threshold selection [12, 22].

2.3. Threshold sensitivity analysis

To evaluate whether the choice of the 95th percentile is justified and robust, a threshold sensitivity analysis was performed on the training dataset. Each threshold was evaluated over 20 independent train–validation splits. For each threshold $p \in [80, 99]$, the angular deviations θ_i were calculated from the training kernel matrix and the points that satisfied $\theta_i > \tau_p$, where $\tau_p = \text{Percentile}_p(\theta)$, were constrained to be orthogonal to the learned regression function, ensuring that their feature-space directions exerted no influence on the predictor.

The results (Figure 2, left) show a consistent decreasing pattern. The threshold (i.e., removing more points) initially improves performance, but after approximately the 95th to 97th percentiles, the validation RMSE stabilizes and reaches its minimum. This demonstrates that the method is not sensitive to the exact choice of percentile, and that the 95th percentile lies within a broad, near-optimal region.

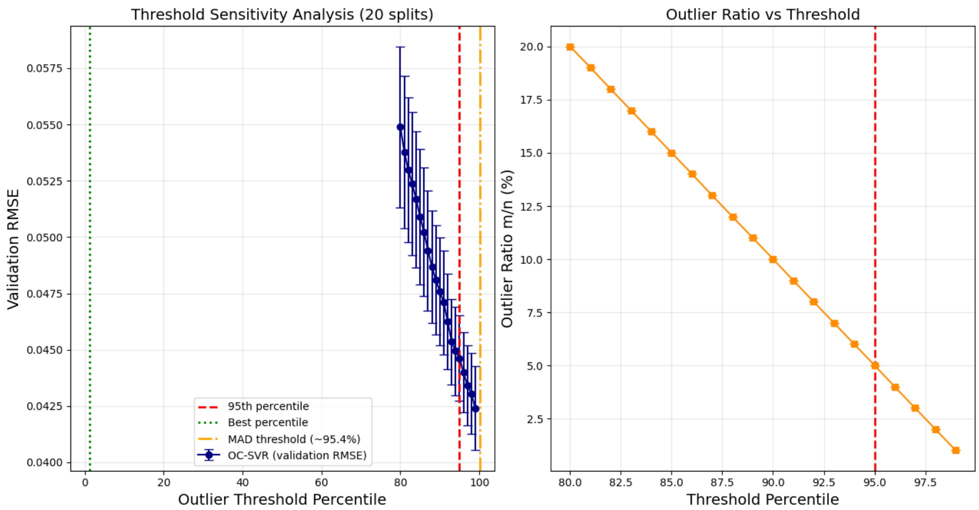

Figure 2. Threshold sensitivity analysis of OC-SVR on the CCPP dataset. (left) Validation RMSE as a function of the angular-deviation percentile threshold, averaged over 20 train–validation splits. Performance stabilizes and reaches its minimum between the 95th and 97th percentiles. (right) Corresponding outlier ratio m/n as a function of the threshold percentile, showing that approximately 5% of samples are identified as outliers at the 95th percentile.

In addition, we compared this to a fully data-driven threshold obtained using the MAD:

$$\tau_{\text{MAD}} = \text{median}(\theta) + 3 \cdot 1.4826 \cdot \text{MAD}(\theta),$$

which corresponds to approximately the 95.4th percentile in our dataset. The performance of this adaptive threshold closely matches that of the fixed 95th percentile, further validating the robustness of the approach.

Figure 2 (right) shows the outlier ratio m/n as a function of the threshold. At the 95th percentile, approximately 5% of the samples are down weighted, aligning with the known contamination level of the CCPP dataset.

In general, these results confirm that the proposed OC-SVR method does not rely on a finely tuned threshold; instead, it exhibits stable performance over a wide range of reasonable values.

2.4. *SVR orthogonal constraint*

The foundation of SVR begins with the primal optimization problem:

$$\min_{w,b,\xi,\xi^*} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{2.7}$$

$$\text{s.t.} \quad y_i - (w^\top \phi(x_i) + b) \le \varepsilon + \xi_i, \quad i = 1, \ldots, n, \tag{2.8}$$

$$(w^\top \phi(x_i) + b) - y_i \le \varepsilon + \xi_i^*, \quad i = 1, \ldots, n, \tag{2.9}$$

$$\xi_i, \xi_i^* \ge 0, \quad i = 1, \ldots, n. \tag{2.10}$$

Here, w and b define the regression function, $\phi(x_i)$ maps the data to an RKHS, ξ_i, ξ_i^* are slack variables, $C > 0$ is a regularization parameter, and ε is the insensitive tube width.

Introducing Lagrange multipliers α_i, α_i^* leads to the SVR dual formulation [2, 17]

$$\max_{\alpha,\alpha^*} \quad -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)K(x_i,x_j) - \varepsilon\sum_{i=1}^{n}(\alpha_i+\alpha_i^*) + \sum_{i=1}^{n} y_i(\alpha_i-\alpha_i^*) \tag{2.11}$$

$$\text{s.t.} \quad \sum_{i=1}^{n}(\alpha_i-\alpha_i^*) = 0, \tag{2.12}$$

$$0 \le \alpha_i,\ \alpha_i^* \le C, \quad i = 1,\ldots,n. \tag{2.13}$$

To reduce the influence of outlier samples, we introduced an orthogonal constraint that collectively separates the regression function from the detected outlier points $x_i \in O$, where $O \subset \{1,\ldots,n\}$ denotes the set of outlier indices. In the primal form, this constraint is expressed as follows:

$$\sum_{i\in O} f(x_i)\,\phi(x_i) = 0_H, \tag{2.14}$$

where $f(x_i) = w^\top\phi(x_i) + b$ and 0_H denotes the zero element in the RKHS H. This equality enforces that the weighted combination of the feature mappings of outlier samples, weighted by their predicted values, vanishes in H. It is important to emphasize that the orthogonal constraint in (2.14) does not enforce pointwise zero predictions for individual outlier samples. Instead, it imposes a collective orthogonality condition in the RKHS, requiring that the regression function be orthogonal to the subspace spanned by the feature mappings of the detected outliers. Only in the special case where the outlier Gram matrix K_{OO} is nonsingular does the constraint reduce to $f(x_i) = 0$ for $i \in O$. In general, the constraint suppresses the directional influence of outliers without forcing their predicted values to vanish individually.

The dual [2] representation of w is written as ,

$$w = \sum_{j=1}^{n}(\alpha_j-\alpha_j^*)\,\phi(x_j), \tag{2.15}$$

and it allows us to express the orthogonal constraint directly in terms of dual variables α_j, α_j^*. This substitution provides the basis for converting the primal vector equality in H into kernel-based scalar constraints suitable for optimization.

To obtain explicit scalar constraints, we took the inner product of both sides of (2.14) with $\phi(x_k)$ for each $k \in O$:

$$\left\langle \sum_{i\in O} f(x_i)\phi(x_i),\ \phi(x_k)\right\rangle_H = \sum_{i\in O} f(x_i)K(x_i,x_k) = 0. \tag{2.16}$$

Stacking these equations for all $k \in O$ gives the matrix form

$$K_{OO}f_O = 0_{|O|}, \tag{2.17}$$

where $K_{OO} \in \mathbb{R}^{|O|\times|O|}$ is the sub-Gram matrix with $[K_{OO}]_{i,k} = K(x_i, x_k)$, and

$$f_O = [f(x_i)]_{i\in O} = K_{XO}^\top\Delta\alpha + b\,\mathbf{1}_{|O|}, \tag{2.18}$$

where $\Delta\alpha_j := \alpha_j - \alpha_j^*$,

$$K_{XO} \in \mathbb{R}^{n\times|O|} \; with \; [K_{XO}]_{j,i} = K(x_j, x_i),$$

and $\mathbf{1}_{|O|}$ is the all-ones vector.

Substituting (2.18) into (2.17) yields the linear constraint:

$$K_{OO}\,(K_{XO}^{\top}\Delta\alpha + b\,\mathbf{1}_{|O|}) = 0_{|O|}. \tag{2.19}$$

If K_{OO} is nonsingular, this simplifies to

$$K_{XO}^{\top}\Delta\alpha + b\,\mathbf{1}_{|O|} = 0_{|O|}, \tag{2.20}$$

which implies $f(x_i) = 0$ for all $i \in O$. If K_{OO} is singular, the condition (2.19) enforces $f_O \in \ker(K_{OO})$, meaning the outlier predictions lie in the null space of K_{OO}.

Integrating the orthogonal constraint into the SVR framework, the SVR dual problem becomes:

$$\begin{aligned}
\max_{\alpha,\alpha^*} \quad & -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta\alpha_i\Delta\alpha_j K(x_i, x_j) \\
& -\varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i\Delta\alpha_i && (2.21)\\
\text{s.t.} \quad & \sum_{i=1}^{n}\Delta\alpha_i = 0, && (2.22)\\
& 0 \le \alpha_i,\, \alpha_i^* \le C, \quad i = 1,\dots,n, && (2.23)\\
& K_{OO}(K_{XO}^{\top}\Delta\alpha + b\,\mathbf{1}_{|O|}) = 0_{|O|}. && (2.24)
\end{aligned}$$

Practical optimization, treatment of the bias term, and stability of the orthogonal constraint

The orthogonal constraint in (2.19)–(2.24) introduces the scalar bias term b, which does not explicitly appear in the SVR dual formulation. To make the optimization structure precise, we explicitly describe the optimization variables and outline two practical solution strategies. We then address numerical stability when K_{OO} is singular and discuss regularization options.

Define the optimization variables $\Delta\alpha_i = \alpha_i - \alpha_i^*, \quad i = 1,\dots,n$, and form the joint optimization vector

$$z = \begin{bmatrix}\Delta\alpha \\ b\end{bmatrix} \in \mathbb{R}^{n+1}.$$

The orthogonal constraint (2.24) can be written compactly as

$$Az = 0, \qquad A = K_{OO}\begin{bmatrix}K_{XO}^{\top} & \mathbf{1}\end{bmatrix} \in \mathbb{R}^{|O|\times(n+1)}. \tag{2.25}$$

(A) Direct quadratic program including b. We treat b as an explicit QP variable. The OC–SVR dual problem becomes:

$$\begin{aligned}
\max_{\alpha,\alpha^*,b} \quad & -\frac{1}{2}\Delta\alpha^{\top}K\Delta\alpha - \varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i\Delta\alpha_i \\
\text{s.t.} \quad & \sum_{i=1}^{n}\Delta\alpha_i = 0, \;\; Az = 0, \;\; 0 \le \alpha_i,\; \alpha_i^* \le C.
\end{aligned} \tag{2.26}$$

This formulation works for any K_{OO} and results in a sparse KKT system of size $(n+|O|+2)\times(n+|O|+2)$.

(B) Eliminating the bias term. If K_{OO} is nonsingular, (2.20) yields

$$K_{XO}^{\top}\Delta\alpha + b\,\mathbf{1} = 0 \quad \Longrightarrow \quad b = -\frac{1}{|O|}\mathbf{1}^{\top}K_{XO}^{\top}\Delta\alpha.$$

Substituting this expression into the dual objective eliminates b entirely, producing a reduced QP in variables $\Delta\alpha$ only. This allows direct use of SVR dual solvers. However, the approach requires K_{OO} to be invertible.

(C) Stability when K_{OO} is singular. Singularity may occur when detected outliers are nearly collinear in the RKHS or when $|O|$ is small. We adopt one of the following stabilizers.

(i) Tikhonov regularization

Replace the constraint matrix by a strictly positive definite one:

$$K_{OO} \quad \rightarrow \quad K_{OO}^{\lambda} = K_{OO} + \lambda I, \qquad \lambda > 0. \tag{2.27}$$

The regularized orthogonal constraint becomes:

$$(K_{OO} + \lambda I)(K_{XO}^{\top}\Delta\alpha + b\mathbf{1}) = 0, \tag{2.28}$$

ensuring numerical stability and well-conditioned KKT systems.

(ii) Relaxed orthogonal constraint

Instead of enforcing exact orthogonality, we allow

$$\left\|K_{OO}(K_{XO}^{\top}\Delta\alpha + b\mathbf{1})\right\|_2^2 \le \eta, \qquad \eta \ge 0, \tag{2.29}$$

or equivalently use a penalty formulation:

$$\min\ \mathcal{L}_{\mathrm{SVR}} + \rho\left\|K_{OO}(K_{XO}^{\top}\Delta\alpha + b\mathbf{1})\right\|_2^2, \qquad \rho > 0. \tag{2.30}$$

This guarantees feasibility even under noisy outlier detection.

Proposition 1. *Stability under regularized orthogonality. Let K be a Mercer kernel and K_{OO} the outlier Gram matrix. For any $\lambda > 0$, the regularized constraint*

$$(K_{OO} + \lambda I)(K_{XO}^{\top}\Delta\alpha + b\mathbf{1}) = 0$$

yields an OC–SVR dual problem that is:

(1) concave in (α, α^*, b),
(2) convex feasible, and
(3) numerically stable, with a full row–rank constraint Jacobian.

Proof. The dual Hessian is $-K$, which is negative semidefinite for Mercer kernels, so the objective is concave. The feasible region is the intersection of affine subspaces and box constraints, hence convex. Since $K_{OO} + \lambda I \succ 0$, the matrix

$$A_\lambda = (K_{OO} + \lambda I)\left[K_{XO}^\top \quad \mathbf{1}\right]$$

has full row rank, giving nonsingular KKT conditions and numerical stability. □

Algorithm 1 OC–SVR Training Algorithm

Require: Training data $\{(x_i, y_i)\}$, kernel K, percentile τ, parameters C, ε

Compute full kernel matrix K
Compute centroid norm $\|c\|$ using K
for $i = 1$ to n **do**
    Compute angular deviation θ_i from (2.4)
end for
Determine threshold $\theta_{\text{th}} = \text{Percentile}_\tau(\{\theta_i\})$
Define outlier set $O = \{i : \theta_i > \theta_{\text{th}}\}$
Form K_{OO} and K_{XO}
if K_{OO} is singular **then**
    $K_{OO} \leftarrow K_{OO} + \lambda I$ (Tikhonov regularization)
end if
Form constraint matrix $A = K_{OO}[K_{XO}^\top \, \mathbf{1}]$
Solve QP:

$$\begin{aligned}\max_{\alpha,\alpha^*,b} \quad & -\tfrac{1}{2}\Delta\alpha^\top K \Delta\alpha - \varepsilon \sum_{i=1}^{n}(\alpha_i + \alpha_i^*) + y^\top \Delta\alpha \\ \text{s.t.} \quad & \mathbf{1}^\top \Delta\alpha = 0, \qquad A[\Delta\alpha; b] = 0, \\ & 0 \le \alpha_i, \alpha_i^* \le C\end{aligned}$$

Output: predictor $f(x) = \sum_{i=1}^{n} \Delta\alpha_i K(x_i, x) + b$

The convexity and feasibility of the OC-SVR dual formulation under the orthogonal constraint can be formally established as follows.

Theorem 1. *Let K be a Mercer kernel such that the Gram matrix*

$$K = [K(x_i, x_j)]_{i,j=1}^{n}$$

is positive semidefinite (PSD). Consider the OC-SVR dual problem:

$$\max_{\alpha,\alpha^*} \; -\tfrac{1}{2}\sum_{i,j} \Delta\alpha_i \Delta\alpha_j K(x_i, x_j) - \varepsilon \sum_i (\alpha_i + \alpha_i^*) + \sum_i y_i \Delta\alpha_i$$

subject to

$$\sum_i \Delta\alpha_i = 0, \quad 0 \le \alpha_i, \alpha_i^* \le C, \quad K_{OO}(K_{XO}^\top \Delta\alpha + b\,\mathbf{1}) = 0.$$

Here $\Delta\alpha_i = \alpha_i - \alpha_i^*$, and K_{XO}, K_{OO} are kernel submatrices corresponding to the full kernel K. Then the dual optimization problem remains a convex quadratic program (concave maximization) with a non-empty feasible set.

Proof. The dual objective contains a quadratic term $-\frac{1}{2}\Delta\alpha^\top K \Delta\alpha$. Because $K \succeq 0$ for a Mercer kernel, the mapping $\Delta\alpha \mapsto -\frac{1}{2}\Delta\alpha^\top K \Delta\alpha$ is concave, and the addition of linear terms preserves concavity. Hence, the objective function remains concave exactly as in the SVR formulation.

The feasible region of the SVR dual is convex since it is defined by:

$$\sum_i \Delta\alpha_i = 0, \quad 0 \leq \alpha_i, \alpha_i^* \leq C,$$

which forms the intersection of an affine hyperplane and box constraints. The orthogonal constraint $K_{OO}(K_{XO}^\top \Delta\alpha + b\mathbf{1}) = 0$ is *linear* in the optimization variables (α, α^*, b). Therefore, adding this constraint preserves convexity, because the intersection of a convex set with an affine subspace remains convex. If K_{OO} is nonsingular, the constraint simplifies to $K_{XO}^\top \Delta\alpha + b\mathbf{1} = 0$, which is an explicitly affine equality.

A feasible solution always exists regardless of whether K_{OO} is singular or nonsingular. For example, choosing $\alpha = \alpha^* = \mathbf{0}, \quad \Delta\alpha = \mathbf{0}, \quad b = 0$ satisfies all standard constraints and trivially satisfies the orthogonal constraint, since $K_{OO}(\mathbf{0}) = \mathbf{0}$. Thus, the feasible set is non-empty.

Hence, the dual objective remains concave under a PSD kernel matrix K; the feasible region remains convex because the orthogonal constraint is affine; and a feasible point exists. Therefore, the OC-SVR dual problem preserves both convexity and feasibility, ensuring that it remains a convex quadratic program solvable by standard QP algorithms. □

Therefore, the modified SVR incorporating the orthogonal constraint remains a convex quadratic program and can be efficiently solved using standard QP solvers.

A key motivation for adopting angular deviation rather than Euclidean distance arises from the geometric distortions induced by kernel mappings. When data are projected into a high-dimensional or infinite-dimensional RKHS, the Euclidean norms of the mapped points tend to inflate and concentrate, a well-known effect associated with distance concentration. Under this phenomenon, Euclidean distance becomes less discriminative, as the majority of points appear to lie at nearly identical distances from each other and from the origin.

Angular deviation, computed from normalized inner products, remains stable under such kernel-induced scaling. Unlike Euclidean distance, the angular measure captures directional information that better reflects the intrinsic structure of high-dimensional data. Outliers typically deviate from the main data manifold not only in magnitude but more critically in direction. By comparing each feature vector with the centroid direction, the angular criterion emphasizes structural misalignment, providing a more robust and informative basis for identifying anomalous samples in RKHS. Consequently, angular deviation is better suited for outlier detection than raw Euclidean distances in kernel spaces.

Most existing robust variants of SVR mitigate the influence of outliers indirectly, for example by modifying the loss function, reweighting samples, or pruning observations deemed unreliable. Although effective to some extent, these heuristics do not explicitly eliminate the geometric influence of outliers within the RKHS. They also often rely on assumptions about noise distribution or penalty shaping, which may not generalize well across different datasets.

The orthogonal constraint proposed in the OC-SVR framework provides a fundamentally different and more principled mechanism for robustness. By enforcing that the regression function be orthogonal to the feature directions of the detected outliers, the model ensures that these samples exert zero net directional influence on the learned predictor. Rather than reducing their weight or altering the optimization landscape, the approach projects the regression function onto the orthogonal complement of the subspace spanned by the outlier feature vectors. This geometric suppression effectively renders the outliers "invisible" to the regression function.

Crucially, the orthogonal constraint preserves the convexity of the dual formulation, distinguishing it from many existing robust SVR approaches. It offers a clear and interpretable geometric rationale: the regression hyperplane is constrained so that its variation cannot align with the directions associated with outliers. This introduces a novel conceptual direction for designing robust kernel-based regression models and provides both theoretical elegance and empirical robustness.

2.5. *Bayesian optimization hyperparameter selection*

Bayesian optimization is an algorithm that efficiently explores the hyperparameter space to find the optimal values. To ensure optimal performance of the OC-SVR, we applied Bayesian optimization to systematically tune key hyperparameters for the kernels used in our simulations.

In our setup, the kernel bandwidth (σ) for the RBF kernel, the regularization term (C) and epsilon (ϵ) for the model were selected through Bayesian optimization. The optimization process was conducted over 30 iterations, evaluating different configurations to identify the most effective set of parameters.

To visualize the optimization process for the optimal hyperparameter search using Bayesian optimization, we plotted the number of iterations against the objective value.

Figure 3 illustrates the progression of the iterations and RMSE over 30 iterations of the Bayesian optimization process. The plot shows how the algorithm gradually searches for the best parameter values by trying different options. The ups and downs in the curve indicate that the algorithm is testing a wide range of parameter combinations to understand how they affect performance. As it gathers more information, the search focuses more on the most promising areas, which leads to lower RMSE values.

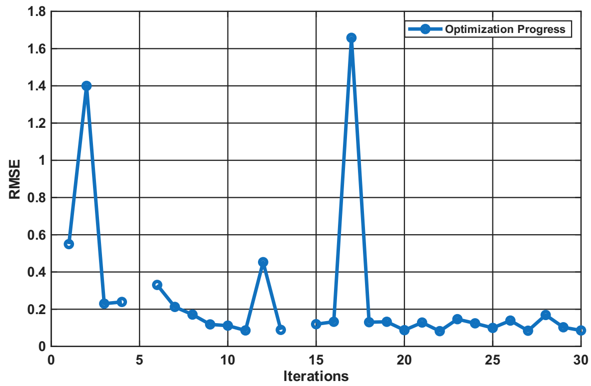

Figure 3. Convergence behavior of Bayesian optimization for OC-SVR hyperparameter tuning on the CCPP dataset. The plot shows the the validation RMSE over 30 optimization iterations, illustrating progressive improvement as the search focuses on promising regions of the hyperparameter space.

In all simulations, the RBF kernel was used due to its universal approximation ability and flexibility in modeling nonlinear relationships. The RBF kernel maps data into an infinite-dimensional feature space, allowing the model to capture complex structures without requiring prior assumptions about the data distribution. Moreover, it has been widely adopted as a benchmark in SVR studies because of its stable performance and minimal hyperparameter requirements compared to other kernels such as polynomial or sigmoid [6, 15, 19]. Since the objective of this work is to evaluate the robustness contribution of the proposed orthogonal constraint rather than to compare kernel types, the RBF kernel provides a consistent and reliable baseline for fair performance evaluation.

2.6. Complexity estimate

The computational complexity of the proposed OC-SVR framework can be divided into two major components. The first is the cost of angular-deviation-based outlier detection, and two is the additional computational and memory overhead introduced by the orthogonal constraint during optimization.

In the outlier detection stage, the kernel matrix $K \in \mathbb{R}^{n\times n}$ is computed, where n is the number of training samples. Since this kernel construction is also required in SVR, it contributes a time and space complexity of $O(n^2)$. The angular deviation θ_i for each data point is calculated from K by averaging across its rows, an additional $O(n^2)$ operation. Consequently, the outlier detection step does not alter the overall asymptotic complexity of the preprocessing phase; it only introduces a small constant overhead.

In the optimization stage, the orthogonal constraint adds m additional linear equality constraints, where m denotes the number of detected outliers. The size of the KKT system expands from $n+1$ SVR to approximately $n+m+1$. For dense interior-point or active-set solvers, this increases the per-iteration computational complexity from $\tilde{O}(n^3)$ to $\tilde{O}((n+m)^3)$, corresponding to a factor of $(1+m/n)^3$. When the outlier proportion is small ($m/n \approx 0.05$), this results in roughly a 15–20% increase in runtime per iteration.

Alternatively, the constraint matrix

$$A = [K_{XO}^{\top}\ \mathbf{1}] \in \mathbb{R}^{m\times(n+1)}$$

can be preprocessed using QR-factorizaton or singular value decomposition to eliminate the constraints prior to optimization. This preprocessing requires a one-time cost of $O((n+1)m^2)$, after which the reduced QP can be solved with the same asymptotic complexity $\tilde{O}(n^3)$ as the SVR.

The memory overhead mainly arises from storing the submatrices $K_{XO} \in \mathbb{R}^{n\times m}$ and $K_{OO} \in \mathbb{R}^{m\times m}$, resulting in $O(nm+m^2)$ additional storage beyond the baseline $O(n^2)$ kernel requirement.

For reference, Table 1 summarizes the computational and memory estimates based on the Combined Cycle Power Plant (CCPP) dataset used in this study ($n = 7654$, $m = 383$, $m/n \approx 5.06\%$).

Table 1. Computational and memory complexity comparison between SVR and OC-SVR.

Component	Standard SVR	OC-SVR	Relative increase
Kernel computation	$O(n^2)$	$O(n^2)$	–
Outlier detection	–	$O(n^2)$	Negligible
QP optimization	$\tilde{O}(n^3)$	$\tilde{O}((n+m)^3)$	$\approx +16\%$
Constraint preprocessing	–	$O((n+1)m^2)$	One-time
Memory usage	$O(n^2)$	$O(n^2+nm+m^2)$	$\approx +5\%$

The angular-deviation outlier detection step does not change the overall order of computational complexity. The orthogonal constraint introduces a modest increase in runtime and memory, scaling linearly with the number of detected outliers. In the CCPP case, this amounts to approximately a 16% increase in computation time and a 5% increase in memory usage, which are small compared to the substantial gains in robustness achieved by the proposed OC-SVR framework.

3. Simulation

In this section, we evaluate the performance of our proposed method, OC-SVR, in comparison with Huber regression, a well-established regression approach recognized for its robustness against noise and outliers. The evaluation was conducted using the dataset of a real-world CCPP obtained from the University of California, Irvine (UCI) machine learning repository [4]. The CCPP dataset from the UCI machine learning repository [4] contains 9568 samples collected from a full load power plant over six years (2006–2011).It includes four features and one target variable.

(1) Independent variables:
- Ambient temperature (AT)
- Exhaust vacuum (V)
- Ambient pressure (AP)
- Relative humidity (RH)

(2) One target variable: Net hourly electrical energy output (PE), measured in megawatts (MW)

The purpose of the simulation is to predict the energy output of the power plant using the OC-SVR method, and the Huber regression. Following the methodology established in Section 2, we first detected outliers for the training samples of the CCPP dataset using the angular deviation method section 2. For outlier detection, the RBF kernel was used to compute the kernel similarities between the training samples in the feature space. The kernel bandwidth(σ) was set to 2.236 ($\gamma = 0.100$), as determined through Bayesian optimization, and a 95th percentile threshold was applied to identify outliers. This process identified 383 outliers, As shown in Figure 4, the cross marker "x" represents the detected outliers, whereas the filled circle marker "o" indicates the normal data points for the CCPP dataset, corresponding to approximately 5.06% of the training set, which were subsequently orthogonalized to enhance model robustness. For training the OC-SVR model,the data set was split into 80% training (7654 samples) and 20% testing (1914 samples). The model parameters were set at the following values: epsilon (ϵ) was fixed at (0.01), the regularization term C was set to 20, and the RBF kernel bandwidth (σ) was determined to be 2.236 ($\gamma = 0.10$). All hyperparameters were selected through Bayesian optimization. The modified SVR with the orthogonal linear constraint Eqs (20–23) was utilized to train the model.

For Huber regression, the Huber loss function was implemented to train the model using the `'robustfit'` function in MATLAB R2025a version. The dataset was partitioned into 80% for training (7,654 samples) and 20% for testing (1,914 samples). All simulations and model implementations were performed in MATLAB R2025a using the Optimization Toolbox. Bayesian optimization was performed using the `bayesopt` function.Additionally,Python 3.13.0 (64-bit, AMD64 build for Windows) was used to generate selected figures.

Table 2. Sample data points from the CCPP dataset.

Index	Temperature	Exhaust Vacuum	Pressure	Humidity	Energy Output
1	14.96	41.76	1024.07	73.17	463.26
2	25.18	62.96	1020.04	59.08	444.37
3	5.11	39.40	1012.16	92.14	488.56
4	20.86	57.32	1010.24	76.64	446.48
5	10.82	37.50	1009.23	96.62	473.90
6	26.27	59.44	1012.23	58.77	443.67
7	15.89	43.96	1014.02	75.24	467.35
8	9.48	44.71	1019.12	66.43	478.42
9	14.64	45.00	1021.78	41.25	475.98
⋮	⋮	⋮	⋮	⋮	⋮
9568	16.12	45.87	1008.15	86.12	457.41

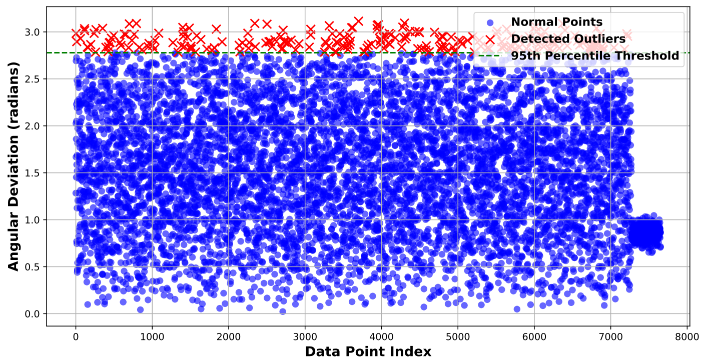

Figure 4. Angular deviation values of training samples in the CCPP dataset. Each point represents a sample indexed along the horizontal axis. Normal samples are shown as circles, while detected outliers (angular deviation above the 95th percentile threshold) are marked with crosses. The horizontal dashed line indicates the selected angular-deviation threshold.

3.1. Results

From the simulation results, it is evident that the newly proposed OC-SVR outperformed the well-established regression variant, Huber regression. Due to the introduction of a linear constraint in the SVR optimization problem, which orthogonalized the impact of outliers, the OC-SVR achieved an RMSE of 0.5324.

Figure 5 illustrates the regression performance on the CCPP dataset, showing that the predicted values are closely aligned with the actual values. This close alignment strongly confirms the effectiveness and robustness of the proposed OC-SVR, particularly in datasets containing conspicuous outliers.

In contrast, the Huber regression, a well-established model known for its robustness to outliers, performed moderately well by mitigating the effects of outliers on the CCPP dataset. It achieved an RMSE of 0.9875. As shown in Figure 6, the predicted values partially align with the actual values with slight deviations. We also evaluated SVR (without any robustness mechanism) on the exact same training/test split of the CCPP dataset, using Bayesian-optimized hyperparameters ($C = 10$, $\varepsilon = 0.3$, $\sigma = 2.236$) and the RBF kernel. SVR produced an RMSE of 2.512 on the test set , substantially higher than the 0.9875 achieved by the Huber regression and the 0.5324 obtained by the proposed OC-SVR. This clearly demonstrates the adverse effect of outliers on conventional SVR and the effectiveness of the orthogonal constraint in mitigating their influence.

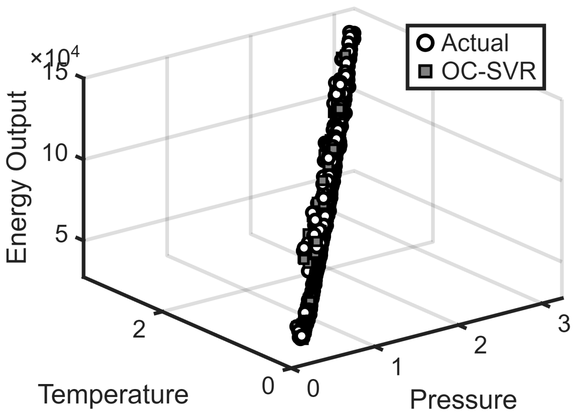

Figure 5. Predicted versus actual energy output values on the test set of the CCPP dataset using the proposed OC-SVR model. The close alignment between predictions and real values indicates strong predictive accuracy and robustness to outliers.

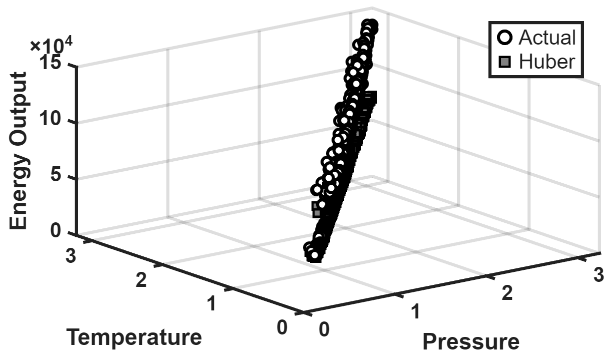

Figure 6. Predicted versus actual energy output values on the test set of the CCPP dataset using Huber regression. Compared with OC-SVR (Figure 5), larger deviations from the real values are observed, reflecting reduced robustness to outliers.

3.2. Additional experiment on synthetic dataset

To further validate the performance of the proposed OC-SVR model, an additional simulation was conducted using synthetic data with controlled outliers. This experiment was designed to assess the model's robustness and predictive accuracy under noisy conditions. A total of 1000 data points were generated from the sinusoidal function $y = \sin(x)$ with additive Gaussian noise ($\sigma = 0.1$). To introduce outliers, 10 points were randomly selected and perturbed with additional noise within the range [2, 3]. Following the methodology described in Section 2, outlier detection was performed using the angular deviation method, where the 95% percentile threshold was applied to identify outliers. The dataset was then partitioned into 80% (800 points) for training and 20% (200 points) for testing. The proposed

OC-SVR model was compared with the SVR method to evaluate relative performance.The RBF kernel was applied to train both the SVR and OC-SVR models. For the SVR method, the parameters were set as $\epsilon = 0.5$, the regularization term $C = 100$, and the kernel bandwidth $\sigma = 1.125$ corresponding to $\gamma = 0.395$. For the proposed OC-SVR, the parameters were $\epsilon = 0.1$, $C = 1.0$, and $\sigma = 1.820$ corresponding to $\gamma = 0.151$. All parameter values for both methods were determined through Bayesian optimization.

As shown in Figure 7, the SVR is noticeably affected by the presence of outliers, resulting in significant deviations from the true sinusoidal function. In contrast, the proposed OC-SVR model effectively mitigates the influence of outliers and provides predictions that closely follow the underlying function. Quantitatively, the OC-SVR achieved a substantially lower RMSE of 0.3210 compared to 1.2630 for the SVR, demonstrating its superior robustness and predictive accuracy on noisy data with outliers.Figure 7 visually compares the predictions of both models on the synthetic dataset.

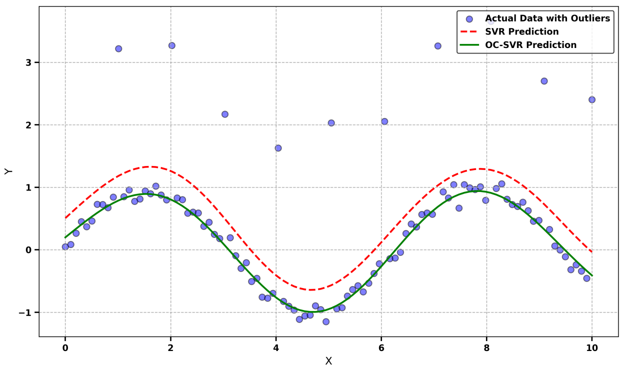

Figure 7. Prediction results on the synthetic sinusoidal dataset with injected outliers. The true underlying function $y = \sin(x)$ is shown together with noisy observations. Predictions from standard SVR and the proposed OC-SVR are overlaid, illustrating that OC-SVR effectively suppresses the influence of outliers and better predicts the underlying function.

Table 3. Summary of experiments across datasets.

Dataset	**Method**	**Kernel**	**RMSE**
CCPP	OC-SVR	RBF	0.5324
CCPP	Huber	–	0.9875
CCPP	SVR	RBF	2.512
Synthetic	OC-SVR	RBF	0.3210
Synthetic	SVR	RBF	1.2630

As observed, the OC-SVR consistently achieved lower RMSE values across both datasets, confirming its robustness against outliers.

4. Limitations of the research

Although the proposed OC-SVR framework demonstrated significant performance improvements, several limitations were identified during this study that warrant consideration and further investigation:

- **Dependence on kernel choice**
  The performance of OC-SVR is inherently dependent on the choice of kernel function. In this study, only the RBF kernel was applied. While effective, this may not generalize optimally to all data types or domains.
- **Fixed outlier threshold selection**
  The angular deviation threshold was determined based on a fixed percentile (95th percentile). Although this approach yielded strong empirical results, it lacks adaptability across datasets with varying noise levels or distributions.
- **Computational complexity in large datasets**
  The inclusion of additional linear constraints and the computation of kernel matrices increase the computational burden, especially for large-scale datasets. This could limit the model's scalability and efficiency in high-dimensional or big data environments.
- **Limited dataset evaluation**
  The validation of OC-SVR was performed using a single real-world dataset (CCPP). While this dataset is representative of real-world regression problems, a broader evaluation across diverse domains, such as finance, and healthcare, would provide stronger generalization evidence.

Addressing these limitations presents opportunities for future extensions of the OC-SVR framework.

5. Conclusions

This study introduced an enhanced SVR framework, termed OC-SVR, which was designed to improve prediction accuracy in the presence of outliers. The proposed approach integrates two key components: (1) a geometric outlier detection mechanism based on angular deviation in the RKHS, and (2) an orthogonal constraint that neutralizes the influence of detected outliers during model training.

In the experiments, the proposed OC-SVR was evaluated on two datasets: the real-world CCPP dataset and a synthetic sinusoidal dataset with deliberately injected outliers. On the CCPP dataset, OC-SVR attained an RMSE of 0.5324, representing improvements of approximately 78.8% and 46.1% over SVR (RMSE = 2.512) and Huber regression (RMSE = 0.9875), respectively. On the synthetic dataset, OC-SVR achieved an RMSE of 0.3210, corresponding to a 74.6% improvement over standard SVR (RMSE = 1.2630) trained and evaluated on identical data. These results confirm that orthogonalizing outlier effects significantly enhances model robustness and generalization.

Overall, the findings highlight that OC-SVR provides a theoretically grounded and scalable regression framework for real-world applications characterized by noise, nonlinearity, and outliers, thereby opening new avenues for robust and accurate prediction.

6. Future work

While the proposed OC-SVR model has demonstrated strong robustness and predictive accuracy, several promising research directions can further extend its applicability and performance:

- **Exploration of alternative kernel functions**
  Future research can investigate the integration of advanced or customized kernel functions,such as

polynomial, sigmoid, Laplacian, or hybrid fuzzy kernels to better capture nonlinear relationships in complex datasets.

- **Adaptive thresholding for outlier detection**
  The current angular deviation threshold was selected based on a fixed percentile approach. Future studies can explore adaptive or dynamic thresholding techniques that adjust automatically based on the statistical properties or distributional shifts of the data. This would improve the model's ability to handle non-stationary or evolving datasets.
- **Multi-objective optimization framework**
  Incorporating multi-objective optimization methods can balance limitations such as prediction accuracy, model sparsity, and robustness to outliers.
- **Extension to other kernel-based learning models**
  The orthogonal constraint framework can be extended beyond SVR to other kernel-based methods such as Support Vector Classification (SVC), Kernel Ridge Regression (KRR), or Gaussian Process Regression (GPR). This would allow the methodology to address a broader range of supervised learning tasks.
- **Large-scale comparison with contemporary robust SVR baselines**
  An extensive empirical evaluation against recent robust SVR variants (e.g., ramp-loss SVR, weighted SVR, HawkEye loss, granular-ball SVR, re-weighted SVR, etc.) on a wider collection of public UCI and real-world datasets is planned for future work. Such a study will further clarify the relative strengths of the proposed orthogonal-constraint mechanism in diverse noisy environments.

Author contributions

Felix Ndudim: Conceptualization, Methodology, Software, Writing-original draft; Thanasak Mouktonglang: Methodology, Validation, Writing-review and editing. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors would like to disclose that OpenAI's ChatGPT was used to assist in improving English grammar and enhancing the clarity of writing in this manuscript. However, the AI tool was not involved in any part of the data analysis or result interpretation.

Acknowledgments

The author gratefully acknowledges the support of the CMU Presidential Scholarship, the Graduate School and the Department of Mathematics, Chiang Mai University, Thailand, for their assistance throughout this research. This research project was partially supported by Fundamental Fund 2025, Chiang Mai University and also Thailand Science Research and Innovation (TSRI) (FRB680102/0162). The authors gratefully acknowledge Chiang Mai University for the research grant and academic support.

Conflict of interest

The authors declare no conflict of interest.

References

1. M. Akhtar, M. Tanveer, M. Arshad, Hawkeye: Advancing robust regression with bounded, smooth, and insensitive loss function, preprint paper, 2024. https://doi.org/10.48550/arXiv.2401.16785
2. N. Cristianini, J. Shawe Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge: Cambridge University Press, 2000.
3. G. De Brabanter, B. De Moor, J. A. K. Suykens, Approximate confidence and prediction intervals for least squares support vector regression, *IEEE Trans. Neural Networks*, **22** (2011), 110–120. https://doi.org/10.1109/TNN.2010.2087769
4. D. Dua, C. Graff, P. Tüfekci, H. Kaya, Combined cycle power plant data set, 2014.
5. U. Gupta, D. Gupta, On regularization based twin support vector regression with huber loss (rhn-tsvr), *Neural Proc. Lett.*, **53** (2021), 459–515. https://doi.org/10.1007/s11063-020-10380-y
6. C. W. Hsu, C. C. Chang, C. J. Lin, A practical guide to support vector classification, Technical report, *Department of Computer Science, National Taiwan University*, 2010. Available from: `https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf`
7. P. J. Huber, Robust estimation of a location parameter, *Ann. Math. Stat.*, **35** (1964), 73–101.
8. Y. Jia, Y. Wang, X. Ma, C. Zhang, Q. Guo, D. Li, A quadratic ν-support vector regression approach for load forecasting, *Complex Intell. Syst.*, **11** (2025), 123. https://doi.org/10.1007/s40747-024-01730-7
9. A. Kocaoğlu, Efficient optimization of a support vector regression model with natural logarithm of the hyperbolic cosine loss function for broader noise distribution, *Appl. Sci.*, **14** (2024), 3641. https://doi.org/10.3390/app14093641
10. H. Luo, S. German Paal, A novel outlier-insensitive local support vector machine for robust data-driven forecasting in engineering, *Eng. Comput.*, **39** (2023), 3671–3689. https://doi.org/10.1007/s00366-022-01781-9
11. R. Muthukrishnan, S. Kalaivani, Robust weighted support vector regression approach for predictive modeling, *Indian J. Sci. Technol.*, **16** (2023), 2287–2296. https://doi.org/10.17485/IJST/v16i30.1180
12. J. Nishiguchi, C. Kaseda, H. Nakayama, M. Arakawa, Y. Yun, Practical approach to outlier detection using support vector regression, In: *Advances in Neuro-Information Processing, Pt I*, 2009, 995–1002. https://doi.org/10.1007/978-3-642-02490-0_121
13. M. Sabzekar, S. M. H. Hasheminejad, Robust regression using support vector regressions, *Chaos Solitons Fract.*, **144** (2021), 110738. https://doi.org/10.1016/j.chaos.2021.110738
14. M. Shi, Y. Xiao, An efficient dual admm for huber regression with fused lasso penalty, *Commun. Stat. Simul. Comput.*, 2025. https://doi.org/10.1080/03610918.2025.2453816

15. A. J. Smola, B. Schölkopf, A tutorial on support vector regression, *Stat. Comput.*, **14** (2004), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

16. B. Sun, X. Liu, Significance support vector regression for image denoising, *Entropy*, **23** (2021), 1262. https://doi.org/10.3390/e23091233

17. H. A. Taha, *Operations Research: An Introduction*, 10 Eds., New York: Macmillan, 2017.

18. T. Uemoto, K. Naito, Support vector regression with penalized likelihood, *Comput. Stat. Data Anal.*, **174** (2022), 107522. https://doi.org/10.1016/j.csda.2022.107522

19. V. Vapnik, *The Nature of Statistical Learning Theory*, Berlin: Springer, 1995.

20. T. Wang, X. Lai, J. Cao, A highly efficient admm-based algorithm for outlier-robust regression with huber loss, *Appl. Intell.*, **54** (2024), 5147–5166. https://doi.org/10.1007/s10489-024-05370-9

21. H. Wei, J. He, Support vector regression model with variant tolerance, *Measur. Control*, **56** (2023), 1135–1143. https://doi.org/10.1177/00202940231180620

22. Y. Zhang, Q. Zhou, B. Du, An algorithm of detecting outliers in svr, *Int. J. Wavel. Multiresol. Inform. Proc.*, **10** (2012), 2012. https://doi.org/10.1142/S0219691312500476

AIMS Press