Check for updates

Partial maximum correntropy regression for robust electrocorticography decoding

Yuanhao Li[1]*, Badong Chen[2], Gang Wang[3], Natsue Yoshimura[4] and Yasuharu Koike[1]

[1]Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Japan, [2]Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China, [3]Key Laboratory of Biomedical Information Engineering of Ministry of Education, Xi'an Jiaotong University, Xi'an, China, [4]School of Computing, Tokyo Institute of Technology, Yokohama, Japan

The *Partial Least Square Regression* (PLSR) method has shown admirable competence for predicting continuous variables from inter-correlated electrocorticography signals in the brain-computer interface. However, PLSR is essentially formulated with the least square criterion, thus, being considerably prone to the performance deterioration caused by the brain recording noises. To address this problem, this study aims to propose a new robust variant for PLSR. To this end, the maximum correntropy criterion (MCC) is utilized to propose a new robust implementation of PLSR, called *Partial Maximum Correntropy Regression* (PMCR). The half-quadratic optimization is utilized to calculate the robust projectors for the dimensionality reduction, and the regression coefficients are optimized by a fixed-point optimization method. The proposed PMCR is evaluated with a synthetic example and a public electrocorticography dataset under three performance indicators. For the synthetic example, PMCR realized better prediction results compared with the other existing methods. PMCR could also abstract valid information with a limited number of decomposition factors in a noisy regression scenario. For the electrocorticography dataset, PMCR achieved superior decoding performance in most cases, and also realized the minimal neurophysiological pattern deterioration with the interference of the noises. The experimental results demonstrate that, the proposed PMCR could outperform the existing methods in a noisy, inter-correlated, and high-dimensional decoding task. PMCR could alleviate the performance degradation caused by the adverse noises and ameliorate the electrocorticography decoding robustness for the brain-computer interface.

KEYWORDS

brain-computer interface, partial least square regression, maximum correntropy, robustness, electrocorticography decoding

1. Introduction

Brain-computer interface (BCI) has been conceived as a promising technology that translates cerebral recordings generated by cortical neurons into appropriate commands for controlling neuroprosthetic devices (Wolpaw et al., 2002). The capability of BCI for repairing or reproducing sensory-motor functions has been increasingly intensified by recent scientific and technological advances (Donoghue, 2002; Mussa-Ivaldi and Miller, 2003; Lebedev and Nicolelis, 2006). The non-invasive recordings, especially electroencephalogram (EEG) and magnetoencephalogram (MEG), are widely exploited to structure BCI systems

due to their ease of use and satisfactory temporal resolution, whereas the non-invasive BCI systems could be limited in their capabilities and customarily require considerable training (Amiri et al., 2013). Invasive single-unit activities and local field potentials commonly provide better decoding performance, which suffer pessimistic long-term stability, however, due to capriciousness in the recorded neuronal-ensembles (Chestek et al., 2007). A sophisticated alternative which exhibits higher signal amplitudes than EEG while presents superior long-term stability compared with invasive modalities, is the semi-invasive electrocorticography (ECoG) (Buzsáki et al., 2012). Numerous studies in recent years have investigated the potentials of ECoG signal for decoding motions (Levine et al., 2000; Leuthardt et al., 2004; Chin et al., 2007; Pistohl et al., 2008; Ball et al., 2009b; Chao et al., 2010; Shimoda et al., 2012). The serviceability of ECoG signal for online practice have also been demonstrated in Leuthardt et al. (2004, 2006), Schalk et al. (2008).

To accomplish the inter-correlated and potentially high-dimensional ECoG decoding tasks, the *partial least square regression* (PLSR) algorithm has been widely utilized to predict continuous variables from ECoG signals as well as various improved versions in the last decade (Chao et al., 2010; Eliseyev et al., 2011, 2012, 2017; Shimoda et al., 2012; Zhao et al., 2012, 2013; Eliseyev and Aksenova, 2016; Foodeh et al., 2020). Chao et al. (2010) successfully predicted the three-dimensional continuous hand trajectories of two monkeys during asynchronous food-reaching tasks from time-frequency features of subdural ECoG signals by PLSR algorithm. They further showed the admirable prediction capability of PLSR in an epidural ECoG study (Shimoda et al., 2012). Recently, different strategies have been investigated to improve the decoding performance of PLSR. For instance, multi-way PLSR algorithms have been proposed as a generalization for tensor analysis in the ECoG decoding tasks (Bro, 1996; Shimoda et al., 2012; Zhao et al., 2013; Eliseyev et al., 2017). Moreover, regularization technique has been used to penalize the objective function with an extra regularization term to achieve desirable prediction (Eliseyev et al., 2012; Eliseyev and Aksenova, 2016; Foodeh et al., 2020). Although the PLSR algorithm was initially developed for econometrics and chemometrics (Wold, 1966), it has emerged as a popular method for neural imaging and decoding (Krishnan et al., 2011; Zhao et al., 2014).

PLSR solves a regression problem primarily with dimensionality reduction on both explanatory matrix (input) and response matrix (output), in which the dimensionality-reduced samples (commonly called as *latent variables*) for respective sets exhibit maximal correlation, thus structuring association from input variables to output variables. Nevertheless, the conventional PLSR and most existing variants are in essence formulated by the least square criterion, which assigns superfluous importance to the deviated noises. On the other hand, although ECoG signal usually exhibits a relatively higher signal-to-noise ratio (SNR) than the non-invasive EEG recording, previous studies have revealed that ECoG is also prone to be contaminated by physiological artifacts with pronounced amplitudes (Otsubo et al., 2008; Ball et al., 2009a). As a result, PLSR could be incompetent for noisy ECoG decoding tasks due to subnormal robustness.

The present study aims to propose a novel robust version for PLSR through introducing the *maximum correntropy criterion* (MCC) to replace the conventional least square criterion, which was proposed in the *information theoretic learning* (ITL) (Principe, 2010), and has achieved the state-of-the-art robust approaches in different tasks, including regression (Liu et al., 2007; Chen and Príncipe, 2012; Feng et al., 2015), classification (Singh et al., 2014; Ren and Yang, 2018), principal component analysis (He et al., 2011), and feature extraction (Dong et al., 2017). Recently, a rudimentary implementation of the MCC in the PLSR algorithm has been investigated in Mou et al. (2018), where MCC was employed in the process of dimensionality reduction. However, the proposed algorithm in Mou et al. (2018) may be limited in some respects. First, except for the MCC-based dimensionality reduction, it remains acquiring the regression relations under the least square criterion. Second, it only considers the dimensionality reduction for the explanatory matrix. Consequently, one has to calculate the regression coefficients separately for each dimension of the response matrix, which means it could be inadequate for multivariate response prediction.

By comparison, the present study aims to realize a more comprehensive implementation of the MCC framework in PLSR. The main contributions of this study are summarized as follows.

1) We reformulate PLSR thoroughly with the MCC framework, that not only the dimensionality reduction, but also the regression relations between the different variables are established by the MCC framework.
2) Both the explanatory matrix (input) and the response matrix (output) are treated with MCC-based dimensionality reduction. As a result, the proposed algorithm is adequate for multivariate response prediction.
3) We utilize Gaussian kernel functions with individual kernel bandwidths for different reconstruction errors and prediction errors. In addition, each kernel bandwidth value could be calculated from the corresponding set of errors directly.

The remainder of this paper is organized as follows. Section 2 introduces the conventional PLSR method as well as the regularized versions. Section 3 gives a brief introduction about MCC and the rudimentary MCC-based PLSR algorithm. Section 4 presents the reformulation of PLSR with the MCC framework, proposing the *partial maximum correntropy regression* (PMCR) algorithm. Section 5 evaluates the proposed method on synthetic and real ECoG datasets, respectively. Some discussions about the proposed method are given in Section 6. Finally, this paper is concluded in Section 7. To facilitate the presentation of this paper, the main notations are listed in Table 1.

2. Partial least square regression

2.1. Conventional PLSR

Consider the data set with the explanatory matrix $\mathbf{X} \in \mathbb{R}^{L \times N}$ and the response matrix $\mathbf{Y} \in \mathbb{R}^{L \times M}$, in which N and M denote the respective numbers of dimension, while L is the

TABLE 1   Main notations.

Notation	Description
L	Number of observations/samples
N	Dimension of explanatory matrix (input)
M	Dimension of response matrix (output)
S	Optimal number of decomposition factors
s	Current index of decomposition factor
$\mathbf{X}$	Original explanatory matrix (input)
$\mathbf{Y}$	Original response matrix (output)
$\hat{\mathbf{Y}}$	Prediction of response matrix
$\mathbf{X}_s$	Residual matrix of $\mathbf{X}$ in s-th factor
$\mathbf{Y}_s$	Residual matrix of $\mathbf{Y}$ in s-th factor
$\mathbf{x}_s^l$	l-th observation in $\mathbf{X}_s$
$\mathbf{y}_s^l$	l-th observation in $\mathbf{Y}_s$
$\mathbf{w}_s$	Dimensionality-reduction projector for $\mathbf{X}_s$
$\mathbf{c}_s$	Dimensionality-reduction projector for $\mathbf{Y}_s$
$\mathbf{t}_s$	Input latent variables in s-th factor
$\mathbf{u}_s$	Output latent variables in s-th factor
$\mathbf{p}_s$	Loading vector in s-th factor
b_s	Regression coefficient between $\mathbf{t}_s$ and $\mathbf{u}_s$
$g_\sigma(\cdot)$	Gaussian kernel function with kernel bandwidth σ

number of observations. PLSR is an iterative regression method which implements dimensionality reduction and decomposition on explanatory and response matrices simultaneously for S iterations, so that they could be expressed by

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T, \; \mathbf{Y} = \mathbf{T}\mathbf{B}\mathbf{C}^T \quad (1)$$

where $\mathbf{T} = [\mathbf{t}_1, .., \mathbf{t}_S] \in \mathbb{R}^{L \times S}$ and $\mathbf{P} = [\mathbf{p}_1, .., \mathbf{p}_S] \in \mathbb{R}^{N \times S}$ are the latent variables and loading vectors for $\mathbf{X}$, respectively. $\mathbf{C} = [\mathbf{c}_1, .., \mathbf{c}_S] \in \mathbb{R}^{M \times S}$ is the loading vectors of $\mathbf{Y}$, and $\mathbf{B} = \text{diag}(b_1, .., b_S) \in \mathbb{R}^{S \times S}$ is a diagonal matrix. For dimensionality reduction, in the s-th iteration with residual matrices $\mathbf{X}_s$ and $\mathbf{Y}_s$, the covariance between the latent variables $\mathbf{t}_s = \mathbf{X}_s\mathbf{w}_s$ and $\mathbf{u}_s = \mathbf{Y}_s\mathbf{c}_s$ are maximized by

$$\max_{\|\mathbf{w}_s\|_2 = \|\mathbf{c}_s\|_2 = 1} \mathbf{t}_s^T\mathbf{u}_s = \mathbf{w}_s^T\mathbf{X}_s^T\mathbf{Y}_s\mathbf{c}_s \quad (2)$$

in which $\mathbf{w}_s \in \mathbb{R}^N$ and $\mathbf{c}_s \in \mathbb{R}^M$ are utilized for dimensionality reduction on $\mathbf{X}_s$ and $\mathbf{Y}_s$, respectively. $\mathbf{u}_s$ is the latent variable for $\mathbf{Y}_s$. $\|\cdot\|_2$ denotes the L_2-norm. After obtaining the latent variables $\mathbf{t}_s$ and $\mathbf{u}_s$, the loading vector $\mathbf{p}_s$ and the relation from $\mathbf{t}_s$ to $\mathbf{u}_s$ with the scalar b_s are founded by the least square criterion

$$\min_{\mathbf{p}_s} \|\mathbf{X}_s - \mathbf{t}_s\mathbf{p}_s^T\|_2^2 \; \Rightarrow \; \mathbf{p}_s = \mathbf{X}_s^T\mathbf{t}_s/(\mathbf{t}_s^T\mathbf{t}_s) \quad (3)$$

$$\min_{b_s} \|\mathbf{u}_s - \mathbf{t}_s b_s\|_2^2 \; \Rightarrow \; b_s = \mathbf{u}_s^T\mathbf{t}_s/(\mathbf{t}_s^T\mathbf{t}_s) \quad (4)$$

The residual matrices are updated by $\mathbf{X}_{s+1} = \mathbf{X}_s - \mathbf{t}_s\mathbf{p}_s^T$ and $\mathbf{Y}_{s+1} = \mathbf{Y}_s - b_s\mathbf{t}_s\mathbf{c}_s^T$. S is usually selected by cross validation. Eventually, the prediction from $\mathbf{X}$ to $\mathbf{Y}$ is structured by

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{H} \quad (5)$$

where $\mathbf{H} = \mathbf{P}^{T+}\mathbf{B}\mathbf{C}^T \in \mathbb{R}^{N \times M}$, and $\mathbf{P}^{T+}$ is the pseudo-inverse of $\mathbf{P}^T$. $\hat{\mathbf{Y}}$ denotes the prediction for $\mathbf{Y}$.

Maximizing the covariance between latent variables Eq. (2) could be rewritten as (Barker and Rayens, 2003).

$$\min_{\|\mathbf{w}_s\|_2 = \|\mathbf{c}_s\|_2 = 1} \sum_{l=1}^{L} \begin{pmatrix} \|\mathbf{x}_s^l - \mathbf{x}_s^l\mathbf{w}_s\mathbf{w}_s^T\|^2 \\ + \|\mathbf{y}_s^l - \mathbf{y}_s^l\mathbf{c}_s\mathbf{c}_s^T\|^2 \\ + \|\mathbf{x}_s^l\mathbf{w}_s - \mathbf{y}_s^l\mathbf{c}_s\|^2 \end{pmatrix} \quad (6)$$

where $\mathbf{x}_s^l$ and $\mathbf{y}_s^l$ denote the l-th samples in $\mathbf{X}_s$ and $\mathbf{Y}_s$, respectively. One can observe that, PLSR employs the least square criterion not only to obtain the regression relations in Eqs. (3, 4), but for the projectors $\mathbf{w}_s$ and $\mathbf{c}_s$ as well. In Eq. (6), the first and second terms are the reconstruction errors for input and output, respectively. The third term denotes the prediction error for the l-th latent variables. Since each step for PLSR is based on the least square criterion, the prediction from input to output could be seriously deteriorated by noises.

2.2. Regularized PLSR

Regularization technique has been popularly employed to ameliorate the decoding performance of the PLSR algorithm. For example, L_1-regularization on the projectors was employed so as to acquire sparse projectors, conducting the feature selection simultaneously (Eliseyev et al., 2012). The authors further extended their study in Eliseyev and Aksenova (2016), in which Sobolev-norm and polynomial penalization were introduced into PLSR algorithm to strengthen the smoothness of the predicted response. Recently, the state-of-the-art regularized PLSR was proposed by utilizing L_2-regularization to find the regression relation between the latent variables $\mathbf{t}_s$ and $\mathbf{u}_s$, so as to reduce the over-fitting risk of each latent variable on the desired response (Foodeh et al., 2020). In particular, for each decomposition factor, the scalar b_s is acquired with an individual regularization parameter λ_s as

$$\min_{b_s} \|\mathbf{u}_s - \mathbf{t}_s b_s\|_2^2 + \lambda_s b_s^2 \; \Rightarrow \; b_s = \mathbf{u}_s^T\mathbf{t}_s/(\mathbf{t}_s^T\mathbf{t}_s + \lambda_s) \quad (7)$$

Experimental results in Foodeh et al. (2020) showed that, the regularization technique in Eq. (7) can achieve better ECoG decoding performance than regularizing the projectors.

Nevertheless, the regularized PLSR variants remain formulated based on the non-robust least square criterion, as a result, being still prone to suffering the performance deterioration caused by the adverse noises.

3. Maximum correntropy criterion

3.1. Maximum correntropy criterion

The correntropy concept was developed in the field of ITL as a generalized correlation function of random processes

(Santamaría et al., 2006), which measures the similarity and interaction between two vectors in a kernel space. Correntropy associates with the information potential of quadratic Renyi's entropy (Liu et al., 2007), where the data's probability density function (PDF) is estimated by the Parzen's window method (Parzen, 1962; Silverman, 1986). The correntropy which evaluates the similarity between two arbitrary variables A and B, is defined by

$$\mathcal{V}(A,B) = E[k(A-B)] \tag{8}$$

in which $k(\cdot)$ is a kernel function satisfying the Mercer's theory and $E[\cdot]$ is the expectation operator. In the practical application, one calculates the correntropy with L observations by the following empirical estimation

$$\hat{\mathcal{V}}(A,B) = \frac{1}{L}\sum_{l=1}^{L}k(a_l - b_l) \tag{9}$$

where the Gaussian kernel function $g_\sigma(x) \triangleq \exp(-x^2/2\sigma^2)$ with kernel bandwidth σ is widely used for the kernel function $k(\cdot)$, thus leading to

$$\hat{\mathcal{V}}(A,B) = \frac{1}{L}\sum_{l=1}^{L}g_\sigma(a_l - b_l) = \frac{1}{L}\sum_{l=1}^{L}\exp(-\frac{(a_l - b_l)^2}{2\sigma^2}) \tag{10}$$

Maximizing the correntropy Eq. (10), called as the *maximum correntropy criterion* (MCC), exhibits numerous advantages. Correntropy is essentially a local similarity measure, which is chiefly determined along $A = B$, i.e. zero-value error. Consequently, the effect of large error caused by adverse noise is alleviated, leading to superior robustness. Additionally, correntropy could extract sufficient information from observations, since it considers all the even moments of errors (Liu et al., 2007). It also relates closely to the *m*-estimation, which can be regarded as a robust formulation of Welsch *m*-estimator (Huber, 2004).

3.2. MCC-PLSR

Recently, a rudimentary MCC-based PLSR variant has been investigated in Mou et al. (2018), named as MCC-PLSR. For a univariate output, according to Mou et al. (2018), the dimensionality reduction Eq. (2) could be rewritten as

$$\max_{\|\mathbf{w}_s\|_2=1} \mathbf{w}_s^T\mathbf{X}_s^T\mathbf{Y}_s\mathbf{Y}_s^T\mathbf{X}_s\mathbf{w}_s \tag{11}$$

which aims to maximize the quadratic covariance. Mou et al. (2018) utilized a similar proposition as in the MCC-based principal component analysis (He et al., 2011), proposing the following objective function

$$\max_{\|\mathbf{w}_s\|_2=1} \sum_{l=1}^{L}g_\sigma\left(\sqrt{\mathbf{y}_s^{lT}\mathbf{x}_s^l\mathbf{x}_s^{lT}\mathbf{y}_s^l - \mathbf{y}_s^{lT}\mathbf{x}_s^l\mathbf{w}_s\mathbf{w}_s^T\mathbf{x}_s^{lT}\mathbf{y}_s^l}\right) \tag{12}$$

from which one can calculate the robust projector $\mathbf{w}_s$. Then, one obtains the latent variables by $\mathbf{t}_s = \mathbf{X}_s\mathbf{w}_s$, and acquires other model parameters similarly as in Eqs. (3-5).

Despite the robust implementation of the projector $\mathbf{w}_s$ in Eq. (12), the above-described MCC-PLSR algorithm could be inadequate for the following reasons. First, except for the calculation of $\mathbf{w}_s$, the other model parameters are still acquired under the least square criterion. Second, dimensionality reduction is not considered for the output matrix. As a result, the prediction performance for multivariate response could be limited. In addition, MCC-PLSR is prone to suffering excessive computation time, since one has to obtain the prediction model $\hat{\mathbf{Y}} = \mathbf{XH}$ for each dimension of the response matrix separately.

4. Partial maximum correntropy regression

In this section, we present a comprehensive reformulation of PLSR with the MCC framework. Compared with the existing MCC-PLSR, our proposed method aims to acquire each model parameter under the MCC. In addition, the generalization for multivariate response prediction is taken into account in this study. The detailed mathematical derivations of the proposed method are given as follows, in which the subscript s denoting the s-th decomposition factor is omitted for the purpose of simplicity.

Substituting the least quadratic reconstruction errors and prediction errors in the conventional PLSR Eq. (6) with the maximum correntropy yields

$$\max_{\|\mathbf{w}\|_2=\|\mathbf{c}\|_2=1} \sum_{l=1}^{L}\left(\begin{array}{c}g_{\sigma_x}(\mathbf{x}^l - \mathbf{x}^l\mathbf{w}\mathbf{w}^T)\\+g_{\sigma_y}(\mathbf{y}^l - \mathbf{y}^l\mathbf{c}\mathbf{c}^T)\\+g_{\sigma_r}(\mathbf{x}^l\mathbf{w} - \mathbf{y}^l\mathbf{c})\end{array}\right) \tag{13}$$

where σ_x, σ_y, and σ_r denote the Gaussian kernel bandwidths for $\mathbf{X}$-reconstruction errors, $\mathbf{Y}$-reconstruction errors, and the prediction errors, respectively.

Then, one can transform the vectors $(\mathbf{x}^l - \mathbf{x}^l\mathbf{w}\mathbf{w}^T)$ and $(\mathbf{y}^l - \mathbf{y}^l\mathbf{c}\mathbf{c}^T)$ into scalars, provided that the two projectors $\mathbf{w}$ and $\mathbf{c}$ are unit-length vectors, i.e. $\mathbf{w}^T\mathbf{w} = \mathbf{c}^T\mathbf{c} = 1$,

$$\begin{array}{c}\sqrt{\|\mathbf{x}^l - \mathbf{x}^l\mathbf{w}\mathbf{w}^T\|^2} = \sqrt{\mathbf{x}^l\mathbf{x}^{lT} - \mathbf{x}^l\mathbf{w}\mathbf{w}^T\mathbf{x}^{lT}}\\\sqrt{\|\mathbf{y}^l - \mathbf{y}^l\mathbf{c}\mathbf{c}^T\|^2} = \sqrt{\mathbf{y}^l\mathbf{y}^{lT} - \mathbf{y}^l\mathbf{c}\mathbf{c}^T\mathbf{y}^{lT}}\end{array} \tag{14}$$

Subsequently, one obtains the following optimization problem to acquire the projectors

$$\max_{\|\mathbf{w}\|_2=\|\mathbf{c}\|_2=1} \sum_{l=1}^{L}\left(\begin{array}{c}g_{\sigma_x}(\sqrt{\mathbf{x}^l\mathbf{x}^{lT} - \mathbf{x}^l\mathbf{w}\mathbf{w}^T\mathbf{x}^{lT}})\\+g_{\sigma_y}(\sqrt{\mathbf{y}^l\mathbf{y}^{lT} - \mathbf{y}^l\mathbf{c}\mathbf{c}^T\mathbf{y}^{lT}})\\+g_{\sigma_r}(\mathbf{x}^l\mathbf{w} - \mathbf{y}^l\mathbf{c})\end{array}\right) \tag{15}$$

After obtaining $\mathbf{w}$ and $\mathbf{c}$, one could calculate the latent variables as in the conventional PLSR by $\mathbf{t} = \mathbf{Xw}$ and $\mathbf{u} = \mathbf{Yc}$. We then calculate the loading vector $\mathbf{p}$ and the regression coefficient b under the MCC by

$$\max_{\mathbf{p}} \sum_{l=1}^{L}g_{\sigma_p}(\mathbf{x}^l - \mathbf{t}^l\mathbf{p}^T) \tag{16}$$

$$\max_{b} \sum_{l=1}^{L}g_{\sigma_b}(\mathbf{u}^l - \mathbf{t}^lb) \tag{17}$$

in which $\mathbf{t}^l$ and $\mathbf{u}^l$ denote the l-th elements for the latent variables $\mathbf{t}$ and $\mathbf{u}$, respectively. σ_p and σ_b denote the corresponding Gaussian kernel bandwidths. The residual matrices are then updated similarly as PLSR.

One repeats such procedures for the optimal number of factors and collects the acquired vectors from each iteration to organize the matrices $\mathbf{T}$, $\mathbf{P}$, $\mathbf{B}$, and $\mathbf{C}$, as in the original PLSR. Ultimately, the predicted response $\hat{\mathbf{Y}}$ can be obtained from $\mathbf{X}$ by the regression relationship Eq. (5). The above-mentioned PLSR variant which is comprehensively reformulated based on the MCC, is named as *partial maximum correntropy regression* (PMCR). In what follows, we discuss in detail about the optimization, convergence analysis, and determination of hyper-parameters with regard to the proposed PMCR algorithm.

4.1. Optimization

Three optimization problems Eqs. (15, 16, 17) need to be addressed in PMCR. We first consider Eq. (15) for the calculation of the projectors $\mathbf{w}$ and $\mathbf{c}$. Based on the *half-quadratic* (HQ) optimization method (Ren and Yang, 2018), Eq. (15) could be rewritten as

$$\max_{\substack{\|\mathbf{w}\|_2 = \\ \|\mathbf{c}\|_2 = 1}} \sum_{l=1}^{L} \begin{pmatrix} \sup\{\alpha_l \frac{\mathbf{x}^l\mathbf{x}^{lT}-\mathbf{x}^l\mathbf{w}\mathbf{w}^T\mathbf{x}^{lT}}{2\sigma_x^2} - \varphi(\alpha_l)\} \\ + \sup\{\beta_l \frac{\mathbf{y}^l\mathbf{y}^{lT}-\mathbf{y}^l\mathbf{c}\mathbf{c}^T\mathbf{y}^{lT}}{2\sigma_y^2} - \varphi(\beta_l)\} \\ + \sup\{\gamma_l \frac{(\mathbf{x}^l\mathbf{w}-\mathbf{y}^l\mathbf{c})^2}{2\sigma_r^2} - \varphi(\gamma_l)\} \end{pmatrix} \quad (18)$$

where $\varphi(\cdot)$ is a convex conjugated function of $g(\cdot)$, and $\{\alpha_l\}_{l=1}^L$, $\{\beta_l\}_{l=1}^L$, and $\{\gamma_l\}_{l=1}^L$ denote three sets of introduced auxiliaries, respectively. Thus, we can conclude that optimizing Eq. (15) is equivalent to updating $(\alpha_l, \beta_l, \gamma_l)$ and $(\mathbf{w}, \mathbf{c})$ alternately by

$$\max_{\substack{\|\mathbf{w}\|_2 = \\ \|\mathbf{c}\|_2 = 1, \\ \alpha_l, \beta_l, \gamma_l}} J \triangleq \sum_{l=1}^{L} \begin{pmatrix} \alpha_l \frac{\mathbf{x}^l\mathbf{x}^{lT}-\mathbf{x}^l\mathbf{w}\mathbf{w}^T\mathbf{x}^{lT}}{2\sigma_x^2} - \varphi(\alpha_l) \\ + \beta_l \frac{\mathbf{y}^l\mathbf{y}^{lT}-\mathbf{y}^l\mathbf{c}\mathbf{c}^T\mathbf{y}^{lT}}{2\sigma_y^2} - \varphi(\beta_l) \\ + \gamma_l \frac{(\mathbf{x}^l\mathbf{w}-\mathbf{y}^l\mathbf{c})^2}{2\sigma_r^2} - \varphi(\gamma_l) \end{pmatrix} \quad (19)$$

Since the HQ optimization is an iterative process, we denote the k-th HQ iteration with the subscript k. First, according to the HQ technique (Ren and Yang, 2018), we update the auxiliaries with the current projectors $(\mathbf{w}_k, \mathbf{c}_k)$ by

$$\alpha_{l,k+1} = -\exp(-\frac{\mathbf{x}^l\mathbf{x}^{lT}-\mathbf{x}^l\mathbf{w}_k\mathbf{w}_k^T\mathbf{x}^{lT}}{2\sigma_x^2})$$

$$\beta_{l,k+1} = -\exp(-\frac{\mathbf{y}^l\mathbf{y}^{lT}-\mathbf{y}^l\mathbf{c}_k\mathbf{c}_k^T\mathbf{y}^{lT}}{2\sigma_y^2}) \quad (20)$$

$$\gamma_{l,k+1} = -\exp(-\frac{(\mathbf{x}^l\mathbf{w}_k-\mathbf{y}^l\mathbf{c}_k)^2}{2\sigma_r^2})$$

$$(l = 1,..,L)$$

Then, to optimize the projectors, we rewrite Eq. (19) by collecting the terms of projectors and omitting the auxiliaries as

$$\max_{\substack{\|\mathbf{w}\|_2 = \\ \|\mathbf{c}\|_2 = 1}} J_p \triangleq \sum_{l=1}^{L} \begin{pmatrix} (\frac{\gamma_l}{2\sigma_r^2} - \frac{\alpha_l}{2\sigma_x^2})\mathbf{x}^l\mathbf{w}\mathbf{w}^T\mathbf{x}^{lT} \\ +(\frac{\gamma_l}{2\sigma_r^2} - \frac{\beta_l}{2\sigma_y^2})\mathbf{y}^l\mathbf{c}\mathbf{c}^T\mathbf{y}^{lT} \\ -\frac{\gamma_l}{\sigma_r^2}\mathbf{x}^l\mathbf{w}\mathbf{c}^T\mathbf{y}^{lT} \end{pmatrix} \quad (21)$$

which is a quadratic optimization issue constrained by nonlinear conditions. To accomplish Eq. (21), there exist enormous solutions in the literature, such as the sequential quadratic programming (SQP) which has been widely utilized for nonlinear programming problems (Fletcher, 2013).

After one obtains the projectors $\mathbf{w}$ and $\mathbf{c}$, the latent variables are computed by $\mathbf{t} = \mathbf{Xw}$ and $\mathbf{u} = \mathbf{Yc}$. Then, Eqs. (16, 17) can be solved by the following iterative fixed-point optimization method with fast convergence (Chen et al., 2015).

$$\mathbf{p} = \mathbf{X}^T\Psi_\mathbf{p}\mathbf{t}/(\mathbf{t}^T\Psi_\mathbf{p}\mathbf{t}) \quad (22)$$

$$b = \mathbf{u}^T\Psi_b\mathbf{t}/(\mathbf{t}^T\Psi_b\mathbf{t}) \quad (23)$$

where $\Psi_\mathbf{p}$ and Ψ_b are $L \times L$ diagonal matrices with the diagonal elements $(\Psi_\mathbf{p})_{l,l} = g_{\sigma_p}(\mathbf{x}^l - \mathbf{t}^l\mathbf{p}^T)$ and $(\Psi_b)_{l,l} = g_{\sigma_b}(\mathbf{u}^l - \mathbf{t}^l b)$, respectively. Since $\Psi_\mathbf{p}$ and Ψ_b are dependent on the current solutions $\mathbf{p}$ and b, the updates in Eqs. (22, 23) are fixed-point equations which will require multiple iterations (Chen et al., 2015). The comprehensive procedures for PMCR are summarized in Algorithm 1.

```
1:  Input: matrices of explanation X and response Y;
    number of factors S; a small positive value ς
2:  Output: prediction model Ŷ = XH
3:  initialize X₁ = X and Y₁ = Y;
4:  for s = 1,2,..,S do
5:      initialize the projectors by the conventional
    PLSR;
6:      initialize converged = FALSE;
7:      repeat
8:          auxiliary-step: update (αₗ,βₗ,γₗ) with (20);
9:          projector-step: update (wₛ,cₛ) with (21);
10:         if the difference of the objective function
    (15) is smaller than ς then
11:             converged = TRUE
12:         end if
13:     until converged == TRUE
14:     compute latent variables tₛ = Xₛwₛ and uₛ = Yₛcₛ;
15:     compute pₛ and bₛ by the fixed-point method
    (22)(23);
16:     update the residual matrices Xₛ₊₁ = Xₛ - tₛpₛᵀ and
    Yₛ₊₁ = Yₛ - bₛtₛcₛᵀ;
17: end for
18: organize the matrices T = [t₁,..,t_S], P = [p₁,..,p_S], B =
    diag(b₁,..,b_S), and C = [c₁,..,c_S];
19: compute H = Pᵀ⁺BCᵀ
```

Algorithm 1. Partial maximum correntropy regression

4.2. Convergence analysis

For the regression relations $\mathbf{p}$ and b, one could find the detailed convergence analysis in Chen et al. (2015). We mainly consider the convergence of the projectors $\mathbf{w}$ and $\mathbf{c}$ in the optimization problem (15). Because correntropy is in nature an m-estimator (Liu et al., 2007), the local optimums of Eq. (15) will be close sufficiently to the global optimum, which has been proved in a recent theoretical study (Loh and Wainwright, 2015). Therefore, we prove that Eq. (15) will converge to a local optimum with the HQ optimization method.

Proposition 1. If we have $J_p(\mathbf{w}_k, \mathbf{c}_k) \leqslant J_p(\mathbf{w}_{k+1}, \mathbf{c}_{k+1})$ by fixing $(\alpha_l, \beta_l, \gamma_l) = (\alpha_{l,k+1}, \beta_{l,k+1}, \gamma_{l,k+1})$, the optimization problem (Eq. 15) will converge to a local optimum.

Proof: The convergence is proved as

$$
\begin{aligned}
& J(\mathbf{w}_k, \mathbf{c}_k, \alpha_{l,k}, \beta_{l,k}, \gamma_{l,k}) \\
& \leqslant J(\mathbf{w}_k, \mathbf{c}_k, \alpha_{l,k+1}, \beta_{l,k+1}, \gamma_{l,k+1}) \\
& \leqslant J(\mathbf{w}_{k+1}, \mathbf{c}_{k+1}, \alpha_{l,k+1}, \beta_{l,k+1}, \gamma_{l,k+1})
\end{aligned}
\tag{24}
$$

in which the first inequality is guaranteed by the HQ mechanism (Ren and Yang, 2018), and the second inequality arises from the assumption of the present proposition.

One can observe that, to guarantee the convergence of Eq. (15), it is unnecessary to attain the strict maximum of Eq. (21) at each projector-step in Algorithm 1. On the contrary, so long as the updated projectors lead to a larger objective function J_p at each projector-step, Eq. (15) will converge to a local optimum. This reveals great convenience in practice, that one only needs a few SQP iterations for projector-step. One could finish the projector-step once confirming the increase on J_p, thus accelerating the convergence.

4.3. Hyper-parameter determination

There exist five Gaussian kernel bandwidths σ_x, σ_y, σ_r, σ_p, and σ_b, respectively, to be determined in practice. In the literature, an effective method to estimate a proper kernel bandwidth for probability density estimation, named as *Silverman's rule*, was proposed in Silverman (1986). Denoting the current set of errors as E with L observations, the kernel bandwidth is computed

$$
\sigma^2 = 1.06 \times \min\{\sigma_E, \frac{R}{1.34}\} \times (L)^{-1/5}
\tag{25}
$$

in which σ_E is the standard deviation of the L errors, and R denotes the interquartile range.

5. Experiments

In this section, we assessed the proposed PMCR algorithm on a synthetic dataset and a real ECoG dataset, respectively, comparing it with the existing PLSR methods. Specifically, we compared PMCR to the following methods: the conventional PLSR, the state-of-the-art regularized PLSR (RPLSR) (Foodeh et al., 2020)

described in Eq. (7), and the rudimentary MCC-PLSR (Mou et al., 2018) described in Section 3.2. For a evenhanded comparison, each algorithm used an identical number of factors, which was selected by the conventional PLSR in five-fold cross-validation. The maximal number of factors was set as 100.

Considering the performance indicators for the evaluation, we used three typical measures in regression tasks: i) Pearson's correlation coefficient (r)

$$
r = \frac{Cov(\hat{\mathbf{Y}}, \mathbf{Y})}{\sqrt{Var(\hat{\mathbf{Y}})Var(\mathbf{Y})}}
\tag{26}
$$

where $Cov(\cdot, \cdot)$ and $Var(\cdot)$ denote the covariance and variance, respectively, and ii) root mean squared error (RMSE) which is computed by

$$
\text{RMSE} = \sqrt{\frac{1}{L} \sum_{l=1}^{L} \|\hat{\mathbf{y}}_l - \mathbf{y}_l\|^2}
\tag{27}
$$

in which $\hat{\mathbf{y}}_l$ and $\mathbf{y}_l$ denote the l-th observations for the prediction $\hat{\mathbf{Y}}$ and the target $\mathbf{Y}$, respectively, and iii) mean absolute error (MAE) which represents the average L_1-norm distance

$$
\text{MAE} = \frac{1}{L} \sum_{l=1}^{L} \|\hat{\mathbf{y}}_l - \mathbf{y}_l\|
\tag{28}
$$

To compare the robustness between different algorithms, only contaminating the training samples by noises with isolating testing data from contamination is an extensively approved and implemented method in the literature, as advised in Zhu and Wu (2004). Accordingly, only the training sample would suffer the adverse contamination in the following experiments.

5.1. Synthetic dataset

5.1.1. Dataset description

First, we considered an inter-correlated, high-dimensional, and noisy synthetic example, in which various PLSR methods were assessed with different levels of contamination. Randomly, we generated 300 i.i.d.[1] latent variables $\mathbf{t} \sim U(0, 1)$ for training, and 300 i.i.d. latent variables $\mathbf{t} \sim U(0, 1)$ for testing, in which U denotes the uniform distribution, and the dimension of $\mathbf{t}$ was set as 20. We generated the hypothesis from the latent variable to the explanatory and response matrices then. Specifically, we randomly generated the transformation matrices with arbitrary values, which were subject to the standard normal distribution. The latent variables $\mathbf{t}$ were multiplied with a 20×500 transformation matrix, resulting in a 300×500 explanatory matrix for input. Similarly, we used a 20×3 transformation matrix to acquire a 300×3 response matrix for output. Accordingly, we predicted the multivariate responses from 500-dimensional explanatory variables with 300 training samples, and evaluated the prediction performance on the other 300 testing samples.

Considering the contamination for the synthetic dataset, we supposed the explanatory matrix to be contaminated, because the

---

1 Independent and identically distributed.

FIGURE 1
Regression performance indicators of the inter-correlated, high-dimensional, and contaminated synthetic dataset under different noise standard deviations with noise levels from 0 to 1.0. **(A)** Noise standard deviation = 30, **(B)** noise standard deviation = 100, and **(C)** noise standard deviation = 300. The performance indicators were acquired from 100 Monte-Carlo repetitive trials and averaged across three dimensions of the output. The proposed PMCR algorithm realized better performance than the existing PLSR algorithms consistently for r, RMSE, and MAE, in particular when the training set was contaminated considerably.

adverse noises mainly happen to the brain recordings, which are usually used as the explanatory in the BCI system. Therefore, a certain proportion (from 0 to 1.0 with a step 0.05) of training samples were randomly selected with equal probability, the inputs of which were then replaced by noises with large amplitude. For the distribution of the noise, we utilized a zero-mean Gaussian distribution with large standard deviation to imitate outliers, where 30, 100, and 300 were used, respectively.

5.1.2. Results

We evaluated the various PLSR methods with 100 Monte-Carlo repetitive trials, and present the results in Figure 1, where the results were averaged across three dimensions of the output. One could observe from Figure 1 that, for all the three different noise standard deviations, the proposed PMCR algorithm achieved superior prediction performance compared with the other existing methods consistently for r, RMSE, and MAE, respectively, in particular when the training set suffered considerable contamination.

The number of factors S plays a vital role in PLSR methods, representing the iteration numbers to decompose the input and output matrices. Since it usually causes a notable effect on the results, additionally, we evaluated the performance with respect to the number of factors for each method. To this end, we utilized the noise standard deviation 100 under three different noise levels, 0.2, 0.5, and 0.8, respectively. The prediction results for each method

are presented in Figure 2 with 100 repetitive trials, with respect to the number of decomposition factors. One could perceive that, not only the proposed PMCR eventually achieved superior regression performance with the optimal number of factors, but also it realized rather commendable performance with a small number of factors. For example, when the noise level was equal to 0.5, the proposed PMCR achieved its optimal performance with no more than 20 factors. By comparison, for the other methods, when the number of factors was larger than 20, their performances remained promoting significantly. One can also observe a similar result in the other two noise levels. This suggests that, PMCR could abstract substantial information with a rather small number of factors from training samples in a noisy regression task.

5.2. ECoG dataset

To further demonstrate the superior robustness of the PMCR algorithm, we evaluated the various PLSR algorithms with the following practical brain decoding task. In this subsection, we used the publicly available Neurotycho ECoG dataset[2] which was initially proposed in Shimoda et al. (2012).

---

2 Available online at http://neurotycho.org/epidural-ecog-food-tracking-task.

FIGURE 2
Regression performance indicators of the synthetic dataset with noise standard deviation being 100 under three different noise levels with the number of factors increasing from 1 to 100. **(A)** Noise level = 0.2, **(B)** noise level = 0.5, and **(C)** noise level = 0.8. The performance indicators were obtained from 100 repetitive trials and averaged across three dimensions of the output. The proposed PMCR algorithm not only acquired better prediction results than the other algorithms ultimately with the optimal number of factors, but also achieved admirable regression performance with a small number of factors.
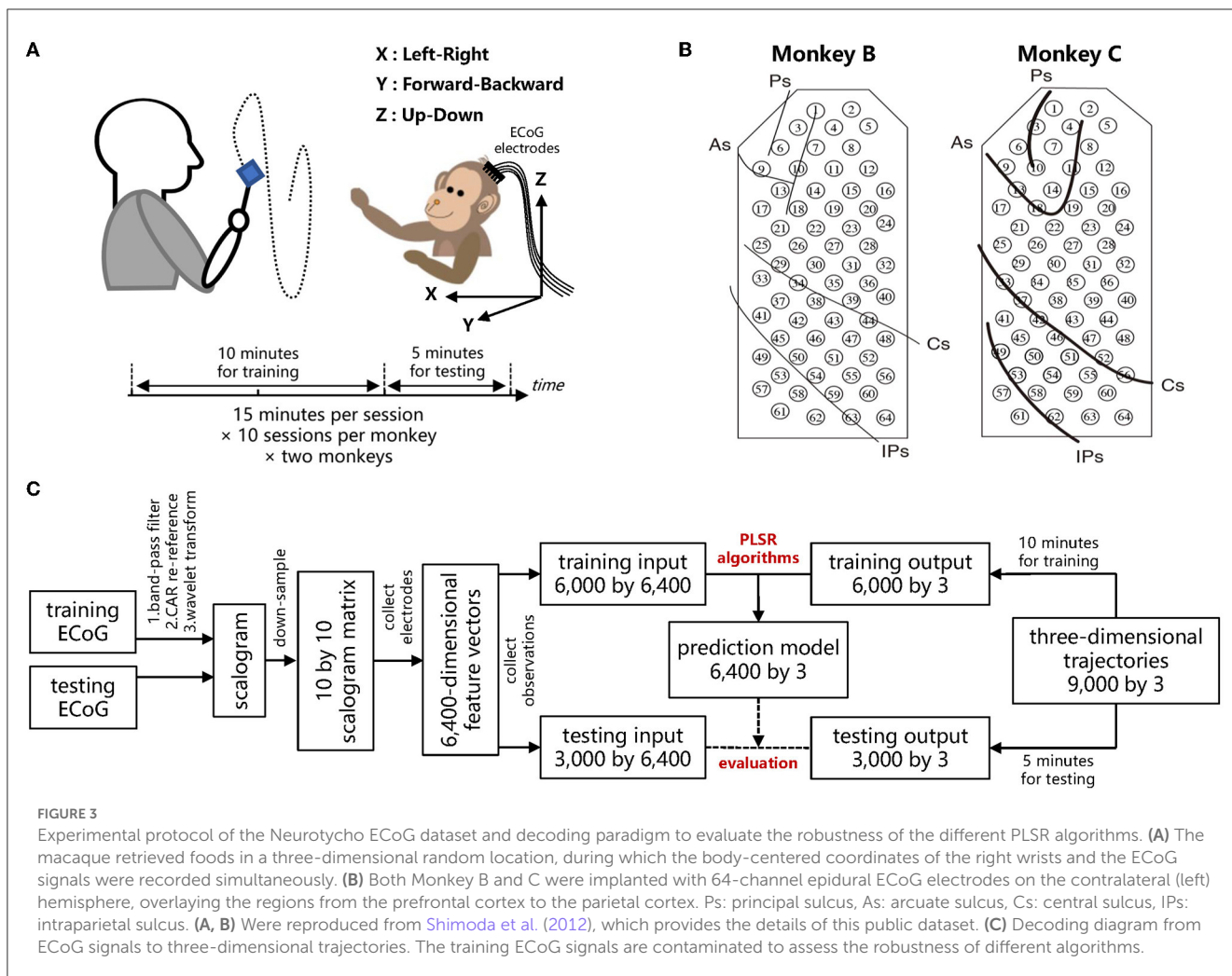
5.2.1. Dataset description

Two Japanese macaques, denoted by Monkey B and C, respectively, were commanded to track foods with the right hands, during which the continuous three-dimensional trajectories of right hands with a sampling rate of 120 Hz were recorded by an optical motion capture instrument. For both Monkey B and C, ten recording sessions were performed, where each recording session lasted 15 minutes. The two macaques were in advance implanted with customized 64-channel ECoG electrodes on the contralateral (left) hemisphere, which covered the regions from the prefrontal cortex to the parietal cortex. ECoG signals were recorded simultaneously during each session with a sampling rate of 1,000 Hz. In accordance with Shimoda et al. (2012), for each recording session, the data of the first ten minutes was used to train a prediction model, while the data of the remaining five minutes was used to evaluate the prediction performance of the trained model. The schemes of the experiments and ECoG electrodes are shown in Figures 3A, B, respectively.

5.2.2. Decoding paradigm

For the feature extraction, we used an identical offline decoding paradigm as in Shimoda et al. (2012). Initially, ECoG signals

were preprocessed with a tenth-order Butterworth bandpass filter with cutoff frequencies from 1 to 400 Hz, and then re-referenced by the common average referencing (CAR) method. The three-dimensional trajectories of the right wrist were down-sampled to 10 Hz, thus, leading to 9,000 samples in one session (10 Hz × 60 sec × 15 min). The three-dimensional position of time t was predicted from the ECoG signals during the previous one second. To extract the features of ECoG signals, we utilized the time-frequency representation. For the time t, the ECoG signals at each electrode from t - 1.1 s to t were processed by Morlet wavelet transformation. Ten center frequencies ranging from 10 to 120 Hz with equal spacing on the logarithmic scale were considered for the wavelet transformation, overlaying the frequency bands which are most relevant to motion tasks (Shimoda et al., 2012). The time-frequency scalogram was then resampled at ten temporal lags with a 0.1 s gap (t - 1 s, t - 0.9 s,..., t - 0.1 s). Thus, the input of each sample exhibited a 6,400-dimensional vector (64 channels × 10 frequencies × 10 temporal lags), and the output was the three-dimensional position of the right hand. Hence, we trained a regression model with 6,000 samples (the first ten minutes) to predict the three-dimensional output from the 6,400-dimensional input, and evaluated the algorithms with other 3,000 testing samples (the remaining five minutes). The illustrative diagrams for ECoG decoding are summarized in Figure 3C.

FIGURE 3
Experimental protocol of the Neurotycho ECoG dataset and decoding paradigm to evaluate the robustness of the different PLSR algorithms. **(A)** The macaque retrieved foods in a three-dimensional random location, during which the body-centered coordinates of the right wrists and the ECoG signals were recorded simultaneously. **(B)** Both Monkey B and C were implanted with 64-channel epidural ECoG electrodes on the contralateral (left) hemisphere, overlaying the regions from the prefrontal cortex to the parietal cortex. Ps: principal sulcus, As: arcuate sulcus, Cs: central sulcus, IPs: intraparietal sulcus. **(A, B)** Were reproduced from Shimoda et al. (2012), which provides the details of this public dataset. **(C)** Decoding diagram from ECoG signals to three-dimensional trajectories. The training ECoG signals are contaminated to assess the robustness of different algorithms.

5.2.3. Contamination

To evaluate the robustness of different algorithms in the practical ECoG decoding task, the ECoG signals were artificially contaminated by outlier to simulate the detrimental artifact. To be specific, we stochastically selected three certain proportions, 0 (no contamination), 10^{-3}, and 10^{-2}, of the training ECoG samplings and corrupted them with outliers which were subject to the zero-mean Gaussian distribution with the variance 50 times that of the signals for the corresponding channel. As stated in Ball et al. (2009a), the blink-related artifacts were remarkably found in ECoG signal that exhibited much larger amplitudes than a normal ECoG recording. Hence, we used the above-mentioned approach to artificially generate adverse artifacts, so as to contaminate the ECoG signals. This method has been widely utilized in the literature to deteriorate the brain signals for evaluating the robustness of different algorithms (Wang et al., 2011; Chen et al., 2018).
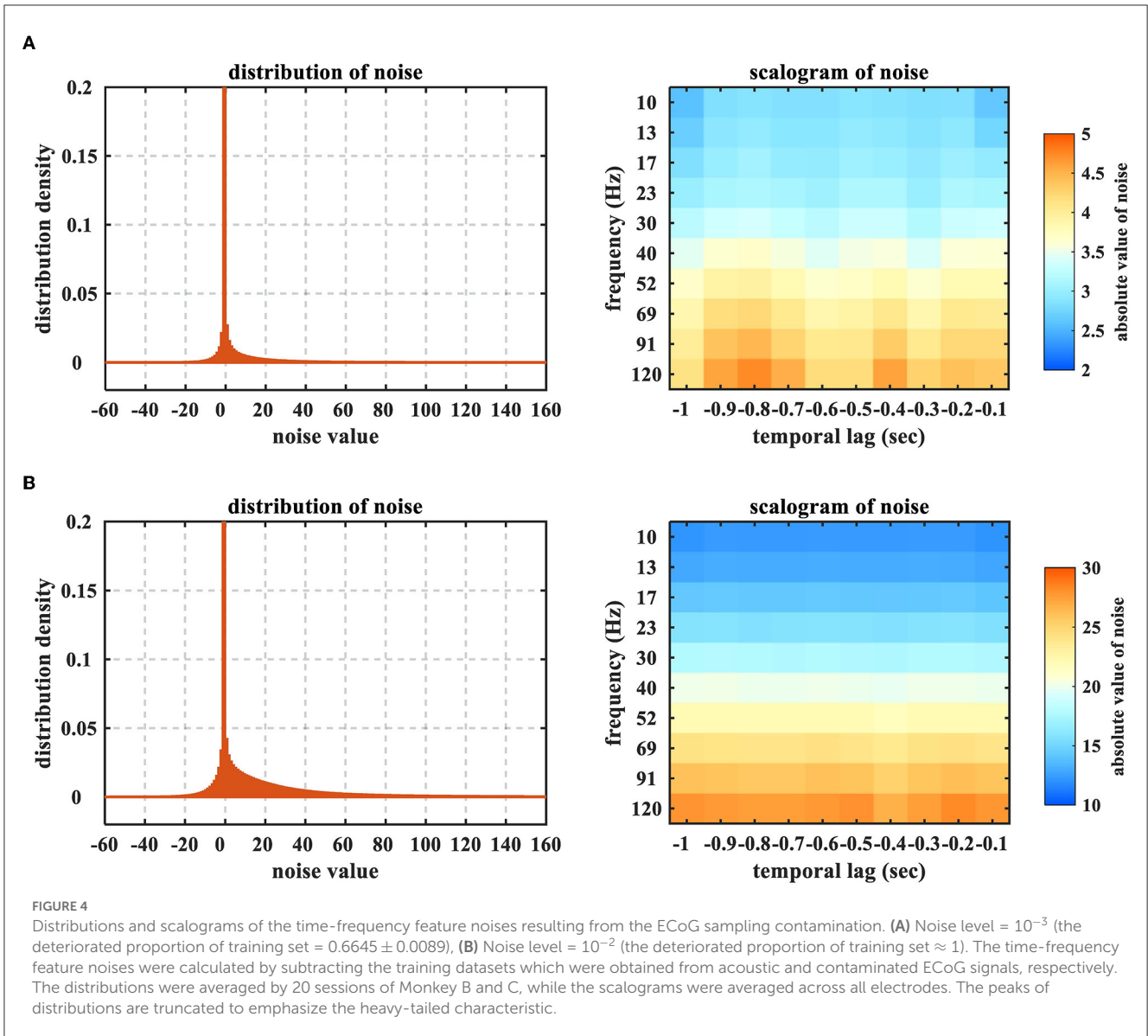
Note that, for this ECoG dataset, the 'Noise Level' signifies the ratio of the contaminated ECoG samplings in the entirety which is different from the ratio of the deteriorated samples in the 6,000 training samples. The ratio of the affected training samples can be evidently larger than the indicated noise level, since one contaminated ECoG sampling could deteriorate several time windows in feature extraction. For example, when the noise level

is denoted as 10^{-3}, the deteriorated proportion of the training set is (0.6645 ± 0.0089). Furthermore, we illustrate how the noise would influence the time-frequency feature in Figure 4. One could obviously perceive the heavy-tailed characteristic on the feature noises, which is in particular intractable for the least square criterion. In addition, the effects of high-frequency band are more prominent, due to the property of impulsive noise.

5.2.4. Spatio-spectro-temporal pattern

Studying how the spatio-spectro-temporal weights in the regression model contribute to the entirety can help investigate the neurophysiological pattern. The element of the trained prediction model $\mathbf{H}$ can be denoted by $h_{\text{ch,freq,temp}}$, which corresponds to the ECoG electrode "ch," the frequency "freq," and the temporal lag "temp." Thus, one could calculate the spatio-spectro-temporal contributions by the ratio between the summation of absolute values of each domain and the summation of absolute values of the entire model

$$W_c(\text{ch}) = \frac{\sum_{\text{freq}} \sum_{\text{temp}} |h_{\text{ch,freq,temp}}|}{\sum_{\text{ch}} \sum_{\text{freq}} \sum_{\text{temp}} |h_{\text{ch,freq,temp}}|} \tag{29}$$

FIGURE 4
Distributions and scalograms of the time-frequency feature noises resulting from the ECoG sampling contamination. **(A)** Noise level = 10^{-3} (the deteriorated proportion of training set = 0.6645 ± 0.0089), **(B)** Noise level = 10^{-2} (the deteriorated proportion of training set ≈ 1). The time-frequency feature noises were calculated by subtracting the training datasets which were obtained from acoustic and contaminated ECoG signals, respectively. The distributions were averaged by 20 sessions of Monkey B and C, while the scalograms were averaged across all electrodes. The peaks of distributions are truncated to emphasize the heavy-tailed characteristic.

$$W_f(\text{freq}) = \frac{\sum_{\text{ch}} \sum_{\text{temp}} |h_{\text{ch,freq,temp}}|}{\sum_{\text{ch}} \sum_{\text{freq}} \sum_{\text{temp}} |h_{\text{ch,freq,temp}}|} \qquad (30)$$

$$W_t(\text{temp}) = \frac{\sum_{\text{ch}} \sum_{\text{freq}} |h_{\text{ch,freq,temp}}|}{\sum_{\text{ch}} \sum_{\text{freq}} \sum_{\text{temp}} |h_{\text{ch,freq,temp}}|} \qquad (31)$$

where $W_c(\text{ch})$, $W_f(\text{freq})$, and $W_t(\text{temp})$ denote the contributions of the ECoG electrode "ch," the frequency "freq," and the temporal lag "temp," respectively.

5.2.5. Results

First, we assessed the different algorithms with the uncontaminated ECoG signals. Accordingly, when the noise level was zero, the average performance indicators were obtained by the acoustic 20 sessions (Monkey B and C). Then we contaminated each session with 5 repetitive trials. Hence, for every noise level, each algorithm was evaluated for 100 times (20 sessions

× 5 repetitive trials). In Table 2, we present the performance indicators for each algorithm with the noise levels 0, 10^{-3}, and 10^{-2}, respectively. In each row of a specific condition, the optimal result is marked in bold. Moreover, the other results are marked with (∗) if there exists statistically significant difference between the current result and the optimal result under each condition. One observes in Table 2 that, the proposed PMCR realized the optimal prediction results consistently, except for the Y-axis under noise level 0. In most cases, PMCR outperformed the other methods with statistically significant difference. One can observe that, when the noise level was 0, PMCR achieved better results than the other algorithms for X-axis and Z-axis. One major reason is, in the acoustic sessions, the motion-related artifacts have been considerably found in the ECoG signals (Shimoda et al., 2012), which further demonstrates the necessity of utilizing PMCR in the practical ECoG decoding tasks.

Furthermore, we studied how the neurophysiological patterns for different algorithms were influenced by the sampling noises.
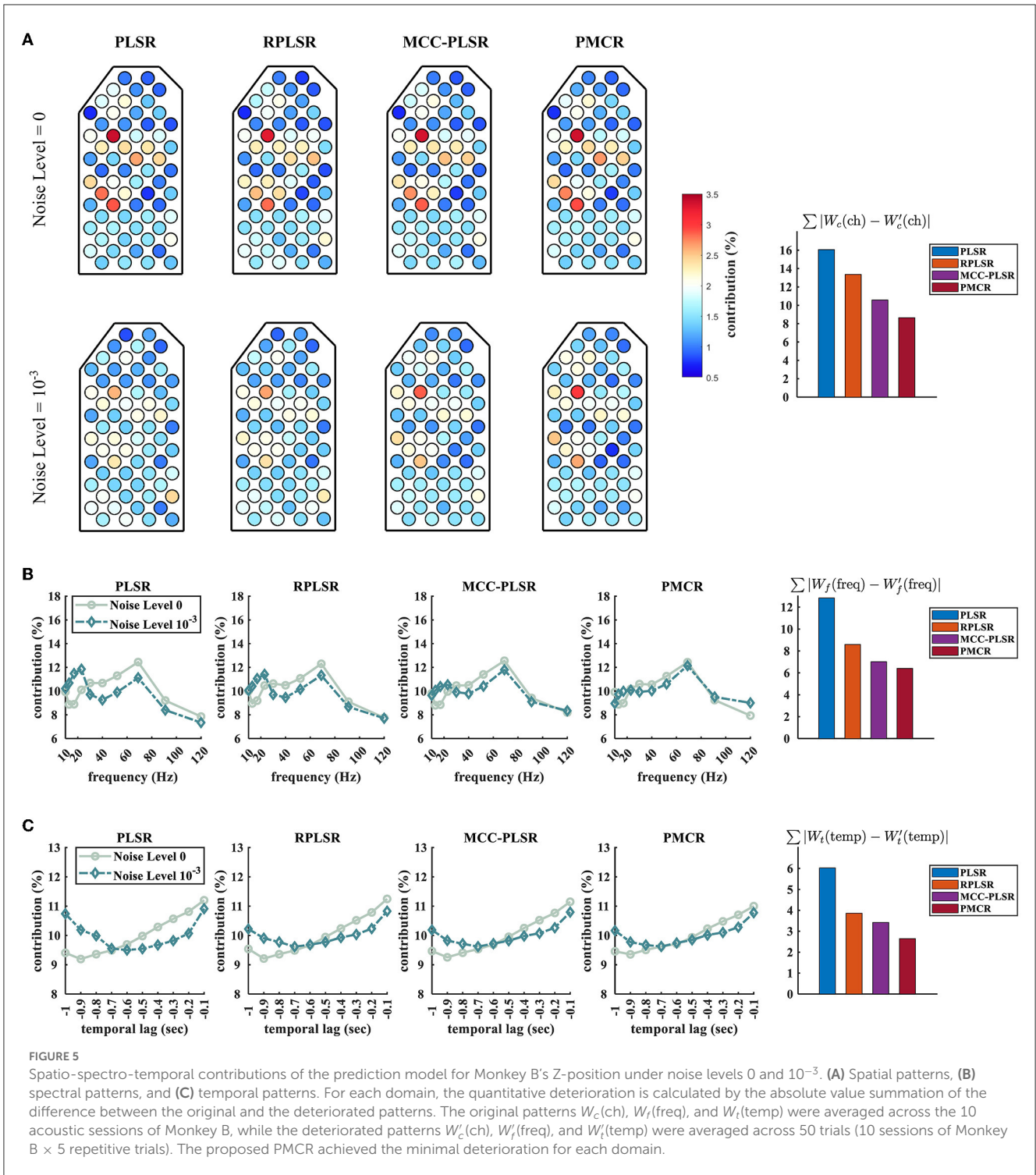
TABLE 2 Performance indicators of each algorithm on the Neurotycho ECoG dataset under three noise levels 0, 10^{-3}, and 10^{-2}, respectively.

X-position						
Algorithm			PLSR	RPLSR	MCC-PLSR	PMCR
Noise Level	0	r	0.4378 ± 0.0933*	0.4550 ± 0.0925*	0.4598 ± 0.0942*	**0.4679 ± 0.0947**
		RMSE	0.9287 ± 0.0810*	0.9037 ± 0.0653*	0.8954 ± 0.0809*	**0.8835 ± 0.0786**
		MAE	0.7026 ± 0.0640*	0.6872 ± 0.0530*	0.6749 ± 0.0628*	**0.6658 ± 0.0651**
	10^{-3}	r	0.3334 ± 0.1165*	0.3558 ± 0.1132*	0.3684 ± 0.1127*	**0.3873 ± 0.1274**
		RMSE	0.9729 ± 0.0652*	0.9543 ± 0.0648*	0.9397 ± 0.0728*	**0.9276 ± 0.0705**
		MAE	0.7291 ± 0.0756*	0.7174 ± 0.0689*	0.7092 ± 0.0786*	**0.6987 ± 0.0759**
	10^{-2}	r	0.1524 ± 0.1399*	0.1713 ± 0.1353*	0.1926 ± 0.1342*	**0.2238 ± 0.1382**
		RMSE	1.0249 ± 0.1105*	1.0022 ± 0.1097*	0.9845 ± 0.1129*	**0.9681 ± 0.1094**
		MAE	0.7655 ± 0.1428*	0.7485 ± 0.1383*	0.7396 ± 0.1392*	**0.7246 ± 0.1397**
Y-position						
Algorithm			PLSR	RPLSR	MCC-PLSR	PMCR
Noise Level	0	r	0.5426 ± 0.1019*	**0.5582 ± 0.1026**	0.5547 ± 0.1017	0.5549 ± 0.1022
		RMSE	0.8483 ± 0.0969*	**0.8198 ± 0.0951**	0.8246 ± 0.0948	0.8233 ± 0.0952
		MAE	0.6487 ± 0.0762*	**0.6304 ± 0.0796**	0.6362 ± 0.0744	0.6358 ± 0.0759
	10^{-3}	r	0.4114 ± 0.1309 *	0.4284 ± 0.1285*	0.4425 ± 0.1302*	**0.4602 ± 0.1296**
		RMSE	0.9188 ± 0.0963 *	0.8962 ± 0.0958*	0.8795 ± 0.0979*	**0.8608 ± 0.1002**
		MAE	0.6960 ± 0.1007*	0.6849 ± 0.1014*	0.6631 ± 0.0983*	**0.6539 ± 0.1021**
	10^{-2}	r	0.2084 ± 0.1514*	0.2206 ± 0.1489*	0.2593 ± 0.1502*	**0.2723 ± 0.1537**
		RMSE	0.9781 ± 0.1143*	0.9542 ± 0.1117*	0.9306 ± 0.1159*	**0.9294 ± 0.1146**
		MAE	0.7354 ± 0.1028*	0.7173 ± 0.1077*	0.7086 ± 0.1105	**0.7043 ± 0.1042**
Z-position						
Algorithm			PLSR	RPLSR	MCC-PLSR	PMCR
Noise Level	0	r	0.6320 ± 0.0324*	0.6395 ± 0.0328*	0.6482 ± 0.0359	**0.6504 ± 0.0372**
		RMSE	0.7968 ± 0.0281*	0.7814 ± 0.0293*	0.7747 ± 0.0296*	**0.7628 ± 0.0275**
		MAE	0.6181 ± 0.0222*	0.6102 ± 0.0280*	0.6055 ± 0.0241	**0.5989 ± 0.0265**
	10^{-3}	r	0.4875 ± 0.0708*	0.4935 ± 0.0701*	0.5158 ± 0.0857*	**0.5259 ± 0.0814**
		RMSE	0.9272 ± 0.0712*	0.9129 ± 0.0682*	0.8958 ± 0.0742*	**0.8834 ± 0.0738**
		MAE	0.6932 ± 0.0800*	0.6894 ± 0.0814*	0.6804 ± 0.0852*	**0.6645 ± 0.0782**
	10^{-2}	r	0.2399 ± 0.1185*	0.2456 ± 0.1173*	0.2615 ± 0.1148*	**0.2803 ± 0.1186**
		RMSE	1.0168 ± 0.0804*	0.9917 ± 0.0785*	0.9605 ± 0.0842*	**0.9485 ± 0.0809**
		MAE	0.7532 ± 0.0883*	0.7429 ± 0.0892*	0.7208 ± 0.0893	**0.7146 ± 0.0887**

The results are given in mean ± deviation, where the optimal results under each condition are marked in bold. The proposed PMCR realized the optimal results consistently, except for the Y-position under the noise level 0. For each result, *is marked if there exists statistically significant difference between the indicated one and the optimal result in the corresponding condition, according to a paired t-test ($p < 0.05$).

We show the differences between the spatial, the spectral, and the temporal contributions which were acquired from the acoustic and the contaminated sessions (under the noise level 10^{-3}), respectively, in Figure 5. The regression model concerning Monkey B's Z-position was used here. We also quantified the effects by computing the summation of the absolute values of the difference between the patterns that were attained from the acoustic and the contaminated sessions, respectively. To be specific, we illustrate $\sum |W_c(\text{ch}) - W_c'(\text{ch})|$, $\sum |W_f(\text{freq}) - W_f'(\text{freq})|$, and $\sum |W_t(\text{temp}) - W_t'(\text{temp})|$ for the spatial, the spectral, and the temporal patterns, respectively. $W_c(\text{ch})$, $W_f(\text{freq})$, and $W_t(\text{temp})$ were obtained by the acoustic sessions, while $W_c'(\text{ch})$, $W_f'(\text{freq})$, and $W_t'(\text{temp})$ were obtained from the contaminated sessions. One can observe from Figure 5 that, the proposed PMCR algorithm realized the minimal deterioration for the pattern of each domain. This further demonstrates the robustness of PMCR in noisy ECoG decoding tasks.

FIGURE 5
Spatio-spectro-temporal contributions of the prediction model for Monkey B's Z-position under noise levels 0 and 10^{-3}. **(A)** Spatial patterns, **(B)** spectral patterns, and **(C)** temporal patterns. For each domain, the quantitative deterioration is calculated by the absolute value summation of the difference between the original and the deteriorated patterns. The original patterns W_c(ch), W_f(freq), and W_t(temp) were averaged across the 10 acoustic sessions of Monkey B, while the deteriorated patterns W_c'(ch), W_f'(freq), and W_t'(temp) were averaged across 50 trials (10 sessions of Monkey B × 5 repetitive trials). The proposed PMCR achieved the minimal deterioration for each domain.

6. Discussion

6.1. Proposed method

In the present study, we aimed to propose a new robust version for PLSR using the MCC framework, which is named as PMCR. Similarly as the existing PLSR methods, the proposed PMCR decomposes the explanatory matrix (input) and the response matrix (output) iteratively for S decomposition factors. The crucial differences of the proposed PMCR are stated in what follows. First, the objective function regarding the projectors $\mathbf{w}_s$ and $\mathbf{c}_s$ in Eq. (15) could be considered as a generalized formulation of the conventional PLSR (Eq. 6), and would be closely related to the calculation in MCC-PLSR (Eq. 12) under specific conditions. As has been proved in Liu et al. (2007), maximizing the correntropy between two variables, when the kernel bandwidth tends to infinity,
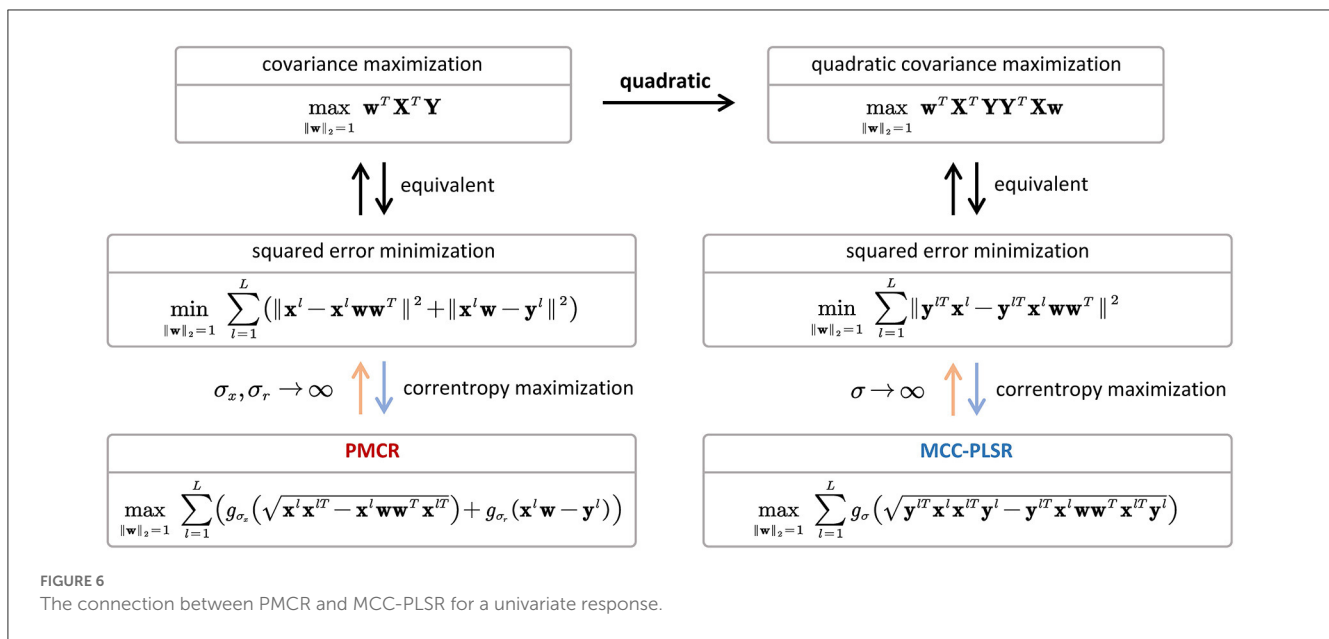
FIGURE 6
The connection between PMCR and MCC-PLSR for a univariate response.

is equal to minimizing their quadratic Euclidean distance. Hence, if we suppose $\sigma_x, \sigma_y, \sigma_r \to \infty$, the projector calculation of PMCR will degenerate to the conventional PLSR. Then, we consider the differences between MCC-PLSR and the proposed PMCR. For a univariate response, the projector $\mathbf{c}$ for dimensionality reduction regarding the response could be ignored. Thus, we can rewrite the dimensionality reduction in PMCR (Eq. 15) as

$$\max_{\|\mathbf{w}\|_2=1} \sum_{l=1}^{L} \left(g_{\sigma_x}(\sqrt{\mathbf{x}^l \mathbf{x}^{lT} - \mathbf{x}^l \mathbf{w} \mathbf{w}^T \mathbf{x}^{lT}}) + g_{\sigma_r}(\mathbf{x}^l \mathbf{w} - \mathbf{y}^l) \right) \quad (32)$$

which could be regarded as a generalized form for the quadratic error minimization (Liu et al., 2007).

$$\min_{\|\mathbf{w}\|_2=1} \sum_{l=1}^{L} \left(\|\mathbf{x}^l - \mathbf{x}^l \mathbf{w} \mathbf{w}^T\|^2 + \|\mathbf{x}^l \mathbf{w} - \mathbf{y}^l\|^2 \right) \Leftrightarrow \max_{\|\mathbf{w}\|_2=1} \mathbf{w}^T \mathbf{X}^T \mathbf{Y}$$
$$(33)$$

which is essentially equal to the conventional PLSR for univariate output. By comparison, MCC-PLSR adopts the MCC framework for the quadratic covariance (Eq. 11), which can be written as Mou et al. (2018)

$$\min_{\|\mathbf{w}\|_2=1} \sum_{l=1}^{L} \|\mathbf{y}^{lT} \mathbf{x}^l - \mathbf{y}^{lT} \mathbf{x}^l \mathbf{w} \mathbf{w}^T\|^2 \Leftrightarrow \max_{\|\mathbf{w}\|_2=1} \mathbf{w}^T \mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X} \mathbf{w} \quad (34)$$

which is the special case of MCC-PLSR when the kernel bandwidth in Eq. (12) tends to infinity. Thus, the connection between PMCR (Eq. 15) and MCC-PLSR (Eq. 12) could be illustrated as in Figure 6. One can observe that, the starting points of PMCR and MCC-PLSR are different. The proposed PMCR begins from the original covariance maximization, whereas MCC-PLSR was proposed from the quadratic covariance. Therefore, we argue that our proposed PMCR is a more rational robust implementation for PLSR. Moreover, note that we give the above discussion under the premise of a univariate output, which is only a special case of degradation for our proposed PMCR. One the other hand,

considering the calculations of the loading vector $\mathbf{p}_s$ and the regression coefficient b_s, the proposed PMCR employs the MCC (Eq. 16, 17), whereas the conventional PLSR and MCC-PLSR utilize the least square criterion. As mentioned above, Eqs. (16, 17) can be also regarded as generalized forms of square error minimization. In summary, the proposed PMCR is more generalized than the conventional PLSR and MCC-PLSR.

In addition, we would like to discuss the advantages and disadvantages of the proposed PMCR algorithm. The essential benefit of utilizing the PMCR algorithm in a noisy ECoG decoding task is the conspicuous robustness with respect to the noises, which was demonstrated with extensive experiments in Section 5. Further, mathematically, the proposed PMCR algorithm is more generalized than the conventional PLSR and MCC-PLSR. As was mentioned above, the conventional PLSR and MCC-PLSR could be regarded as special cases of the proposed PMCR under specific conditions concerning the kernel bandwidths. In particular, compared with MCC-PLSR, the proposed PMCR takes into account the dimensionality reduction for the response matrix. As a result, PMCR could realize better prediction performance for multivariate response. Moreover, PMCR could be further implemented with regularization techniques and extended to the multi-way scenario, which would be discussed in the following subsections. However, PMCR might suffer the performance degradation resulting from inadequate kernel bandwidths that are calculated by the *Silverman's rule* (Eq. 25). Although the experimental results in this study verified empirically that, the proposed PMCR could perform efficiently with the kernel bandwidths acquired by Eq. (25), it may be difficult to guarantee that the *Silverman's rule* can always provide adequate kernel bandwidths. Hence, we would like to investigate a better way to determine the kernel bandwidths with solid theoretical guarantees in our future works. In addition, our proposed PMCR is effective to deal with outliers, while it may be inadequate for multi-modal-distributed noise because MCC utilizes only one kernel function for each reconstruction error. To address this issue, it is promising to use *minimum error entropy*

(MEE) to reformulate PLSR, another popular learning criterion in ITL (Principe, 2010). MEE employs multiple kernel functions for each reconstruction error, so that it can realize satisfactory robustness with respect to multi-modal-distributed noise, which has realized robust neural decoding algorithms (Chen et al., 2018; Li et al., 2021).

6.2. PMCR with regularization

One should additionally note that, the PMCR was proposed by reformulating the conventional PLSR algorithm with using the robust MCC, instead of the mediocre least square criterion. Hence, the proposed PMCR exhibits the supplementary potential for further performance improvements with regularization techniques, as well as in the existing regularized PLSR methods. For example, L_1-regularization could be utilized in Eq. (15) to encourage sparse and robust projectors. In addition, if one requires better smoothness on the predicted output, polynomial or Sobolev-norm penalization could be utilized in PMCR. Moreover, L_2-regularization could be utilized for Eq. (17) to decrease the over-fitting risk considering the regression scalar b_s, similarly as Eq. (7) (Foodeh et al., 2020). In the literature, MCC-based algorithms with regularization have been widely investigated. For instance, a robust version of sparse representation classifier (SRC) for face recognition was developed by employing L_1-regularization on the MCC-based SRC objective function (He et al., 2010).

6.3. Extension to multi-way application

The multi-way PLSR establishes the regression relationship between tensor variables with dimensionality reduction by tensor factorization technique. In the literature, the multi-way PLSR was usually reported to achieve better decoding capability than the generic PLSR algorithm in the brain decoding task, where the spatio-spectro-temporal feature is organized with the tensor form. Essentially, the multi-way PLSR decomposes the input and output under the least square criterion by minimizing the Frobenius-norm (Kolda and Bader, 2009). Therefore, the multi-way PLSR is prone to the performance deterioration caused by noises as well.

The proposed PMCR method treats the regression problem of matrix, i.e. two-way variable. Extending the PMCR algorithm to multi-way application could probably improve the prediction performance further, which would be investigated in our future works. Promisingly, MCC has been demonstrated effective for tensor variable analysis in a recent study (Zhang et al., 2016).

7. Conclusion

This paper proposed a new robust variant for the PLSR algorithm by reformulating the non-robust least square criterion

with the sophisticated MCC framework. The proposed robust objective functions can be effectively optimized by half-quadratic and fixed-point optimization methods. Extensive experimental results with the synthetic dataset and Neurotycho epidural ECoG dataset demonstrate that, the proposed PMCR can outperform the existing PLSR algorithms, revealing promising robustness for high-dimensional and noisy ECoG decoding tasks.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

YL and BC contributed to conceptualization of the study and developed the algorithm derivation. YL and GW operated the experiments and analyzed the results. YL wrote the first draft of the manuscript. NY and YK directed the study and guided the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Amiri, S., Fazel-Rezai, R., and Asadpour, V. (2013). "A review of hybrid brain-computer interface systems," in *Advances in Human-Computer Interaction 2013*. doi: 10.1155/2013/187024

Ball, T., Kern, M., Mutschler, I., Aertsen, A., and Schulze-Bonhage, A. (2009a). Signal quality of simultaneously recorded invasive and non-invasive EEG. *Neuroimage* 46, 708–716. doi: 10.1016/j.neuroimage.2009.02.028

Ball, T., Schulze-Bonhage, A., Aertsen, A., and Mehring, C. (2009b). Differential representation of arm movement direction in relation to cortical anatomy and function. *J. Neur. Eng.* 6, 016006. doi: 10.1088/1741-2560/6/1/016006

Barker, M., and Rayens, W. (2003). Partial least squares for discrimination. *J. Chemometr.* 17, 166–173. doi: 10.1002/cem.785

Bro, R. (1996). Multiway calibration: multilinear PLS. *J. Chemometr.* 10, 47–61. doi: 10.1002/(SICI)1099-128X(199601)10:147::AID-CEM4003.0.CO;2-C

Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents–EEG, ECOG, LFP, and spikes. *Nat. Rev. Neurosci.* 13, 407–420. doi: 10.1038/nrn3241

Chao, Z. C., Nagasaka, Y., and Fujii, N. (2010). Long-term asynchronous decoding of arm motion using electrocorticographic signals in monkey. *Front. Neuroeng.* 3, 3. doi: 10.3389/fneng.2010.00003

Chen, B., Li, Y., Dong, J., Lu, N., and Qin, J. (2018). Common spatial patterns based on the quantized minimum error entropy criterion. *IEEE Trans. Syst. Man Cybern.* 50, 4557–4568. doi: 10.1109/TSMC.2018.2855106

Chen, B., and Príncipe, J. C. (2012). Maximum correntropy estimation is a smoothed map estimation. *IEEE Signal Proc. Lett.* 19, 491–494. doi: 10.1109/LSP.2012.2204435

Chen, B., Wang, J., Zhao, H., Zheng, N., and Principe, J. C. (2015). Convergence of a fixed-point algorithm under maximum correntropy criterion. *IEEE Signal Proces. Lett.* 22, 1723–1727. doi: 10.1109/LSP.2015.2428713

Chestek, C. A., Batista, A. P., Santhanam, G., Byron, M. Y., Afshar, A., Cunningham, J. P., et al. (2007). Single-neuron stability during repeated reaching in macaque premotor cortex. *J. Neurosci.* 27, 10742–10750. doi: 10.1523/JNEUROSCI.0959-07.2007

Chin, C. M., Popovic, M. R., Thrasher, A., Cameron, T., Lozano, A., and Chen, R. (2007). Identification of arm movements using correlation of electrocorticographic spectral components and kinematic recordings. *J. Neur. Eng.* 4, 146. doi: 10.1088/1741-2560/4/2/014

Dong, J., Chen, B., Lu, N., Wang, H., and Zheng, N. (2017). "Correntropy induced metric based common spatial patterns," in *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)* (IEEE) 1–6. doi: 10.1109/MLSP.2017.8168132

Donoghue, J. P. (2002). Connecting cortex to machines: recent advances in brain interfaces. *Nat. Neurosci.* 5, 1085–1088. doi: 10.1038/nn947

Eliseyev, A., and Aksenova, T. (2016). Penalized multi-way partial least squares for smooth trajectory decoding from electrocorticographic (ecog) recording. *PLoS ONE* 11, e0154878. doi: 10.1371/journal.pone.0154878

Eliseyev, A., Auboiroux, V., Costecalde, T., Langar, L., Charvet, G., Mestais, C., et al. (2017). Recursive exponentially weighted n-way partial least squares regression with recursive-validation of hyper-parameters in brain-computer interface applications. *Sci. Rep.* 7, 1–15. doi: 10.1038/s41598-017-16579-9

Eliseyev, A., Moro, C., Costecalde, T., Torres, N., Gharbi, S., Mestais, C., et al. (2011). Iterative n-way partial least squares for a binary self-paced brain-computer interface in freely moving animals. *J. Neur. Eng.* 8, 046012. doi: 10.1088/1741-2560/8/4/046012

Eliseyev, A., Moro, C., Faber, J., Wyss, A., Torres, N., Mestais, C., et al. (2012). L1-penalized n-way pls for subset of electrodes selection in bci experiments. *J. Neur. Eng.* 9, 045010. doi: 10.1088/1741-2560/9/4/045010

Feng, Y., Huang, X., Shi, L., Yang, Y., and Suykens, J. A. (2015). Learning with the maximum correntropy criterion induced losses for regression. *J. Mach. Learn. Res.*, 16, 993–1034.

Fletcher, R. (2013). *Practical Methods of Optimization*. New York: John Wiley Sons.

Foodeh, R., Ebadollahi, S., and Daliri, M. R. (2020). Regularized partial least square regression for continuous decoding in brain-computer interfaces. *Neuroinformatics* 18, 465–477. doi: 10.1007/s12021-020-09455-x

He, R., Hu, B.-G., Zheng, W.-S., and Kong, X.-W. (2011). Robust principal component analysis based on maximum correntropy criterion. *IEEE Trans. Image Proc.* 20, 1485–1494. doi: 10.1109/TIP.2010.2103949

He, R., Zheng, W.-S., and Hu, B.-G. (2010). Maximum correntropy criterion for robust face recognition. *IEEE Trans. Patt. Analy. Mach. Intell.* 33, 1561–1576. doi: 10.1109/TPAMI.2010.220

Huber, P. J. (2004). *Robust Statistics*, vol. 523. Oxford: John Wiley Sons.

Kolda, T. G., and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* 51, 455–500. doi: 10.1137/07070111X

Krishnan, A., Williams, L. J., McIntosh, A. R., and Abdi, H. (2011). Partial least squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage* 56, 455–475. doi: 10.1016/j.neuroimage.2010.07.034

Lebedev, M. A., and Nicolelis, M. A. (2006). Brain-machine interfaces: past, present and future. *TRENDS Neurosci.* 29, 536–546. doi: 10.1016/j.tins.2006.07.004

Leuthardt, E. C., Miller, K. J., Schalk, G., Rao, R. P., and Ojemann, J. G. (2006). Electrocorticography-based brain computer interface-the seattle experience. *IEEE Trans. Neur. Syst. Rehabilit. Eng.* 14, 194–198. doi: 10.1109/TNSRE.2006.875536

Leuthardt, E. C., Schalk, G., Wolpaw, J. R., Ojemann, J. G., and Moran, D. W. (2004). A brain-computer interface using electrocorticographic signals in humans. *J. Neur. Eng.* 1, 63. doi: 10.1088/1741-2560/1/2/001

Levine, S. P., Huggins, J. E., BeMent, S. L., Kushwaha, R. K., Schuh, L. A., Rohde, M. M., et al. (2000). A direct brain interface based on event-related potentials. *IEEE Trans. Rehabil. Eng.* 8, 180–185. doi: 10.1109/86.847809

Li, Y., Chen, B., Yoshimura, N., and Koike, Y. (2021). Restricted minimum error entropy criterion for robust classification. *IEEE Trans. Neur. Netw. Learn. Syst.* 33, 6599–6612. doi: 10.1109/TNNLS.2021.3082571

Liu, W., Pokharel, P. P., and Principe, J. C. (2007). Correntropy: Properties and applications in non-gaussian signal processing. *IEEE Trans. Signal Process.* 55, 5286–5298. doi: 10.1109/TSP.2007.896065

Loh, P.-L., and Wainwright, M. J. (2015). Regularized m-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *J. Mach. Learn. Res.* 16, 559–616.

Mou, Y., Zhou, L., Chen, W., Fan, J., and Zhao, X. (2018). Maximum correntropy criterion partial least squares. *Optik* 165, 137–147. doi: 10.1016/j.ijleo.2017.12.126

Mussa-Ivaldi, F. A., and Miller, L. E. (2003). Brain-machine interfaces: computational demands and clinical needs meet basic neuroscience. *TRENDS Neurosci.* 26, 329–334. doi: 10.1016/S0166-2236(03)00121-8

Otsubo, H., Ochi, A., Imai, K., Akiyama, T., Fujimoto, A., Go, C., et al. (2008). High-frequency oscillations of ictal muscle activity and epileptogenic discharges on intracranial eeg in a temporal lobe epilepsy patient. *Clin. Neurophysiol.* 119, 862–868. doi: 10.1016/j.clinph.2007.12.014

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Mathem. Stat.* 33, 1065–1076. doi: 10.1214/aoms/1177704472

Pistohl, T., Ball, T., Schulze-Bonhage, A., Aertsen, A., and Mehring, C. (2008). Prediction of arm movement trajectories from ecog-recordings in humans. *J. Neurosci. Methods* 167, 105–114. doi: 10.1016/j.jneumeth.2007.10.001

Principe, J. C. (2010). *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*. New York: Springer Science Business Media. doi: 10.1007/978-1-4419-1570-2

Ren, Z., and Yang, L. (2018). Correntropy-based robust extreme learning machine for classification. *Neurocomputing* 313, 74–84. doi: 10.1016/j.neucom.2018.05.100

Santamaría, I., Pokharel, P. P., and Principe, J. C. (2006). Generalized correlation function: definition, properties, and application to blind equalization. *IEEE Trans. Signal Proces.* 54, 2187–2197. doi: 10.1109/TSP.2006.872524

Schalk, G., Miller, K. J., Anderson, N. R., Wilson, J. A., Smyth, M. D., Ojemann, J. G., et al. (2008). Two-dimensional movement control using electrocorticographic signals in humans. *J. Neur. Eng.* 5, 75. doi: 10.1088/1741-2560/5/1/008

Shimoda, K., Nagasaka, Y., Chao, Z. C., and Fujii, N. (2012). Decoding continuous three-dimensional hand trajectories from epidural electrocorticographic signals in japanese macaques. *J. Neur. Eng.* 9, 036015. doi: 10.1088/1741-2560/9/3/036015

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis, volume 26*. New York: CRC press.

Singh, A., Pokharel, R., and Principe, J. (2014). The c-loss function for pattern classification. *Patt. Recogn.* 47, 441–453. doi: 10.1016/j.patcog.2013.07.017

Wang, H., Tang, Q., and Zheng, W. (2011). L1-norm-based common spatial patterns. *IEEE Trans. Biomed. Eng.* 59, 653–662. doi: 10.1109/TBME.2011.2177523

Wold, H. (1966). "Estimation of principal components and related models by iterative least squares," in *Multivariate analysis*, eds. P. R. Krishnajah (NewYork: Academic Press) 391–420.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3

Zhang, M., Gao, Y., Sun, C., La Salle, J., and Liang, J. (2016). "Robust tensor factorization using maximum correntropy criterion," in *2016 23rd International Conference on Pattern Recognition (ICPR)* (IEEE) 4184–4189.

Zhao, Q., Caiafa, C. F., Mandic, D. P., Chao, Z. C., Nagasaka, Y., Fujii, N., et al. (2012). Higher order partial least squares (hopls): A generalized multilinear regression method. *IEEE Trans. Patt. Analy. Mach. Intell.* 35, 1660–1673. doi: 10.1109/TPAMI.2012.254

Zhao, Q., Zhang, L., and Cichocki, A. (2014). Multilinear and nonlinear generalizations of partial least squares: an overview of recent advances. *Wiley Interdisc. Rev.* 4, 104–115. doi: 10.1002/widm.1120

Zhao, Q., Zhou, G., Adalı, T., Zhang, L., and Cichocki, A. (2013). "Kernel-based tensor partial least squares for reconstruction of limb movements," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE) 3577–3581. doi: 10.1109/ICASSP.2013.6638324

Zhu, X., and Wu, X. (2004). Class noise vs. attribute noise: A quantitative study. *Artif. Intell. Rev.* 22, 177–210. doi: 10.1007/s10462-004-0751-8