EEG-Based User Reaction Time Estimation Using Riemannian Geometry Features

Dongrui Wu, *Senior Member, IEEE*, Brent J. Lance, *Senior Member, IEEE*, Vernon J. Lawhern, *Member, IEEE*, Stephen Gordon, Tzyy-Ping Jung, *Fellow, IEEE*, and Chin-Teng Lin, *Fellow, IEEE*

***Abstract*— Riemannian geometry has been successfully used in many brain–computer interface (BCI) classification problems and demonstrated superior performance. In this paper, for the first time, it is applied to BCI regression problems, an important category of BCI applications. More specifically, we propose a new feature extraction approach for electroencephalogram (EEG)-based BCI regression problems: a spatial filter is first used to increase the signal quality of the EEG trials and also to reduce the dimensionality of the covariance matrices, and then Riemannian tangent space features are extracted. We validate the performance of the proposed approach in reaction time estimation from EEG signals measured in a large-scale sustained-attention psychomotor vigilance task, and show that compared with the traditional powerband features, the tangent space features can reduce the root mean square estimation error by 4.30%–8.30%, and increase the estimation correlation coefficient by 6.59%–11.13%.**

***Index Terms*— Brain-computer interface, EEG, reaction time estimation, Riemannian geometry, spatial filtering.**

I. Introduction

BRAIN-COMPUTER interfaces (BCIs) can use brain signals such as the scalp electroencephalogram (EEG) to enable people to communicate or control external devices [23], [36]. Thus, they can help people with devastating neuromuscular disorders such as amyotrophic lateral sclerosis, brainstem stroke, cerebral palsy, and spinal cord injury [53]. However, there are still many challenges in their transition from laboratory settings to real-life applications, including the reliability and convenience of the sensing hardware [29], and the availability of high-performance and robust algorithms for signal analysis and interpretation [24], [33], [34], [47]. This paper focuses on the latter, particularly, feature extraction for EEG-based BCIs.

Riemannian geometry (RG) [3], [10], [27], [41], [45] is a very useful mathematical tool in machine learning and signal/image processing, due to its utility in generating smooth manifolds from intrinsically nonlinear data spaces. Recently it has also been introduced into the BCI community and demonstrated superior performance in a number of applications [5]–[8], [14], [25], [28], [33], [39], [48], [59].

For example, Li *et al.* [28] used RG of the EEG power spectral density matrices for sleep pattern classification. They also proposed a closed-form weighting matrix for the power spectral density matrices to minimize the distance between similar features and to maximize the distance between dissimilar features, and demonstrated better performance than the Euclidian distance and the Kullback-Leibler distance. Barachant et al. [5] proposed two RG approaches for motor imagery classification. The first uses the spatial covariance matrices of the EEG signal as features and RG to directly classify them in the manifold of symmetric and positive definite (SPD) matrices. The second maps the covariance matrices onto the Riemannian tangent space, which is a Euclidean space, and then performs variable selection and classification. They achieved comparable or better performance than a multiclass Common Spatial Pattern (CSP) plus Linear Discriminant Analysis (LDA) approach. Congedo *et al.* [14] further used RG to build calibrationless BCI systems for applications based on event-related potentials, sensorimotor (mu) rhythms, and steady-state evoked potential. It outperformed several state-of-the-art approaches, including xDAWN, stepwise LDA, CSP+LDA, and blind source separation plus logistic regression. Barachant [7] also proposed a spatial filter to increase the signal to signal-plus-noise ratio of magnetoencephalography (MEG) signals before constructing a special form of a covariance matrix for RG feature extraction, and a k-means clustering like unsupervised learning algorithm

Manuscript received October 18, 2016; revised March 19, 2017; accepted April 24, 2017. Date of publication April 28, 2017; date of current version November 6, 2017. This work was supported in part by the U.S. Army Research Laboratory under Agreement W911NF-10-2-0022 and Agreement W911NF-10-D-0002/TO 0023 and in part by the Australian Research Council under Grant DP150101645. *(Corresponding author: Dongrui Wu.)*

D. Wu is with DataNova, Clifton Park, NY 12065 USA (e-mail: drwu09@gmail.com).

B. J. Lance is with the Human Research and Engineering Directorate, U.S. Army Research Laboratory, Aberdeen Proving Ground, Aberdeen, MD 21005 USA (e-mail: brent.j.lance.civ@mail.mil).

V. J. Lawhern is with the Human Research and Engineering Directorate, U.S. Army Research Laboratory, Aberdeen Proving Ground, Aberdeen, MD 21005 USA, and also with the Department of Computer Science, The University of Texas at San Antonio, San Antonio, TX 78249 USA (e-mail: vernon.j.lawhern.civ@mail.mil).

S. Gordon is with DCS Corporation, Alexandria, VA 22310 USA (e-mail: sgordon@dcscorp.com).

T.-P. Jung is with the Swartz Center for Computational Neuroscience, University of California at San Diego, La Jolla, CA 92093 USA, and also with the Center for Advanced Neurological Engineering, University of California at San Diego, La Jolla, CA 92093 USA (e-mail: jung@sccn.ucsd.edu).

C.-T. Lin is with the Centre for Artificial Intelligence, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: chin-teng.lin@uts.edu.au).

Digital Object Identifier 10.1109/TNSRE.2017.2699784

1534-4320 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

in the Riemannian manifold to improve the offline classification performance. This approach outperformed 266 other approaches and won the Kaggle "DecMeg2014 – Decoding the Human Brain" competition,[1] which aimed to predict visual stimuli from MEG recordings of human brain activity. Kalunga et al. [25] proposed an online classification approach in the Riemannian space and showed that it outperformed Canonical Correlation Analysis in Steady-State Visually Evoked Potential classification. Yger *et al.* [59] empirically compared several covariance matrix averaging methods for EEG signal classification. They showed that RG for averaging covariance matrices improved performances for small dimensional problems, but as the dimensionality of the covariance matrix increased, RG became less efficient. Lotte [33] also proposed a framework to combine transfer learning, ensemble learning, and RG for calibration time reduction, which outperformed CSP+LDA. The Riemannian distance was used in regularization to emphasize auxiliary users whose covariance matrices are close to the target user. Navarro-Sune et al. [39] proposed a BCI to automatically detect patient-ventilator disharmony from EEG signals. RG of EEG covariance matrices was used in semi-supervised learning for effective classification of respiratory state, and it outperformed the Euclidean distance. Waytowich et al. [48] proposed an approach to integrate RG with transfer learning and spectral meta-learner [40], an offline ensemble fusion approach, for user-independent BCI, and demonstrated in single-trial event-related potential classification that it can significantly outperform existing calibration-free techniques and traditional within-subject calibration techniques when limited data is available.

All above approaches focused on EEG classification problems in BCI, whereas BCI regression problems have been largely overlooked. In theory a regression problem is equivalent to a classification problem with infinitely many classes, and hence the output has much finer granularity than a traditional two-class or multi-class classification problem, which provides richer information in decision making. There are at least two types of BCI regression problems in the literature and practice. The first type is behavioral or cognitive status prediction, e.g., estimating the continuous value of a driver's drowsiness from the EEG [30]–[32], [49], [54], [56]–[58], and estimating a subject's response speed in a psychomotor vigilance task (PVT) from the EEG [55]. The second type is direct control applications, e.g., controlling the movement of a mouse cursor using BCI [13], [21], [35], [51], [52], and controlling the continuous movement of a hand in the 3D space using EEG [12].

Once the EEG signal is acquired, the regression problem involves three steps: 1) signal processing to increase the signal-to-noise ratio. Frequency domain filters, such as band pass filters and notch filters [12], [13], and spatial filters, such as independent component analysis [30] and CSP [55], are frequently used here. 2) feature extraction to construct meaningful predictors, e.g., standardized difference of the EEG voltage [12], [13], and EEG power band features [54], [55], [57], [58]. 3) regression algorithms to estimate the continuous output, e.g., ordinary linear regression [12], [13], ridge regression [54], LASSO [55], k-nearest neighbors (kNN) [55], fuzzy neural networks [32], transfer learning [49], [56], active learning [57], etc.

In this paper, we apply RG and tangent space features to supervised BCI regression problems. To overcome the limitation pointed out by Yger *et al.* [59], i.e., RG is less efficient when the dimensionality of the covariance matrix is large, we adopt an approach similar to what Barachant used in [7]: we first use a spatial filter proposed in [55] to reduce the dimensionality of the covariance matrices and also to increase the EEG signal quality, and then extract the RG features in the Riemannian tangent space. We validate the performance of the proposed approach in reaction time (RT) estimation from EEG signals measured in a large-scale sustained-attention PVT [16], which collected 143 sessions of data from 17 subjects in a 5-month period. To our knowledge, this is the first time that RG has been used in BCI regression problems.

The remainder of this paper is organized as follows: Section II describes the spatial filter we proposed earlier for supervised BCI regression problems. Section III introduces RG and the tangent space features for BCI regression problems. Section IV describes the experimental setup, RT and EEG data preprocessing techniques, and the procedure to evaluate the performances of different feature extraction methods. Section V presents the results of the comparative studies. Section VI provides parameter sensitivity analysis and additional discussions. Finally, Section VII draws conclusions and outlines a future research direction.

II. Spatial Filtering for Supervised BCI Regression Problems

Recently we [55] proposed two spatial filters for supervised BCI regression problems, which were extended from the common spatial pattern (CSP) algorithm for supervised classification problems. They have similar performance and computational cost. One of them, CSP for regression - one versus the rest (CSPR-OVR), is briefly introduced in this section, as the RG features are better extracted from the spatially filtered EEG data than the raw EEG data.

Let $\mathbf{X}_n \in \mathbb{R}^{C\times S}$ $(n = 1, \ldots, N)$ denote the nth EEG trial in the training data, where C is the number of channels and S the number of time samples. We assume that the mean of each channel measurement has been removed, which is usually performed by band-pass filtering. Let $y_n \in \mathbb{R}$ be the corresponding RT of the nth trial. CSPR-OVR first constructs K fuzzy sets [60], which partition the training samples into K fuzzy classes. To do that, it partitions the interval $[0, 100]$ into $K + 1$ equal intervals, and denotes the partition points as $\{p_k\}_{k=1,\ldots,K}$. It is easy to obtain that

$$p_k = \frac{100 \cdot k}{K+1}, \qquad k = 1, \ldots, K \tag{1}$$

For each p_k, CSPR-OVR then finds the corresponding p_k percentile value of all training y_n and denotes it as P_k. Next we define K fuzzy classes from them, as shown in Fig. 1.

[1] https://www.kaggle.com/c/decoding-the-human-brain.

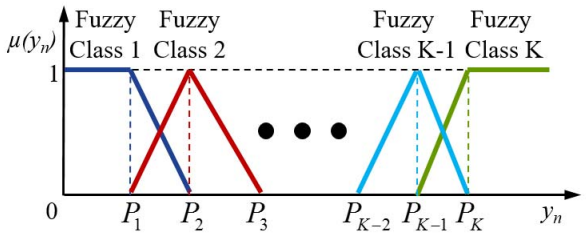


Fig. 1. The K fuzzy classes for y_n.

Then, for each fuzzy class, CSPR-OVR computes its mean spatial covariance matrix as:

$$\bar{\Sigma}_k = \frac{\sum_{n=1}^{N} \mu_k(y_n)\mathbf{X}_n\mathbf{X}_n^T}{\sum_{n=1}^{N} \mu_k(y_n)}, \qquad k = 1, \ldots, K \tag{2}$$

where $\mu_k(y_n)$ is the membership degree of y_n in Fuzzy Class k.

Next CSPR-OVR designs a spatial filtering matrix $\mathbf{W}_k^* \in \mathbb{R}^{C\times F}$, where F is the number of individual vector filters, to maximize the variance difference between Fuzzy Class k and the rest, i.e.,

$$\mathbf{W}_k^* = \arg\max_{\mathbf{W}\in\mathbb{R}^{C\times F}} \frac{\mathrm{Tr}(\mathbf{W}^T \bar{\Sigma}_k \mathbf{W})}{\mathrm{Tr}[\mathbf{W}^T(\sum_{i\neq k} \bar{\Sigma}_i)\mathbf{W}]} \tag{3}$$

where $\mathrm{Tr}(\cdot)$ is the trace of a matrix. (3) is a generalized Rayleigh quotient [22], and the solution $\mathbf{W}_k^*$ is the concatenation of the F eigenvectors associated with the F largest eigenvalues of the matrix $(\sum_{i\neq k} \bar{\Sigma}_i)^{-1}\bar{\Sigma}_k$.

The final spatial filtering matrix $\mathbf{W}^* \in \mathbb{R}^{C\times KF}$ is the concatenation of all $\mathbf{W}_k^*$, i.e.,

$$\mathbf{W}^* = [\mathbf{W}_1^*, \quad \ldots, \quad \mathbf{W}_K^*] \tag{4}$$

and the spatially filtered trial for $\mathbf{X}_n$ is:

$$\mathbf{X}_n' = \mathbf{W}^{*T}\mathbf{X}_n, \quad n = 1, \ldots, N. \tag{5}$$

In summary, the complete CSPR-OVR algorithm for supervised BCI regression problems is shown in Algorithm 1.

Algorithm 1 The CSPR-OVR Spatial Filter for Supervised BCI Regression Problems [55]

Input: EEG training examples $(\mathbf{X}_n, y_n)$, where $\mathbf{X}_n \in \mathbb{R}^{C\times S}$, $n = 1, \ldots, N$;
K, the number of fuzzy classes for y_n;
F, the number of spatial filters for each fuzzy class.
Output: Spatially filtered EEG trials $\mathbf{X}_n' \in \mathbb{R}^{KF\times S}$.
Band-pass filter each $\mathbf{X}_n$ to remove the mean of each channel;
Compute $\{p_k\}_{k=1,\ldots,K}$ in (1);
Compute the corresponding percentile values $\{P_k\}_{k=1,\ldots,K}$ for y_n;
Construct the K fuzzy classes as shown in Fig. 1;
Compute $\bar{\Sigma}_k$ by (2);
Compute $\mathbf{W}_k^*$ by (3);
Construct $\mathbf{W}^*$ by (4);
Return $\mathbf{X}_n'$ by (5)

III. RG and the Tangent Space Features

This section introduces the basics of RG, and an approach to extract the Riemannian tangent space features.

A. Riemannian Geometry

The RG approach for BCI works on the covariance matrices of EEG trials, which are symmetric positive-definite and form a differentiable Riemannian manifold $\mathcal{M}$ [20] with dimensionality $R(R+1)/2$, where R is the number of rows (columns) of the covariance matrices. As a result, we need to use Riemannian metrics, instead of the traditional Euclidean metrics, which are more appropriate for flat spaces of vectors. Particularly, we are interested in the distance measure between two covariance matrices, as many machine learning methods rely on such distances.

The *Riemannian distance* $\delta(\bar{\Sigma}, \Sigma_n)$ between two covariance matrices $\bar{\Sigma} \in \mathbb{R}^{R\times R}$ and $\Sigma_n \in \mathbb{R}^{R\times R}$, called the *geodesic*, is the minimum length of a curve connecting them on the manifold $\mathcal{M}$. It can be computed as [4], [37]:

$$\delta(\bar{\Sigma}, \Sigma_n) = \left\|\log\left(\bar{\Sigma}^{-1}\Sigma_n\right)\right\|_F = \left[\sum_{r=1}^{R} \log^2 \lambda_r\right]^{\frac{1}{2}} \tag{6}$$

where the subscript $_F$ denotes the Frobenius norm, and λ_r, $r = 1, \ldots, R$, are the real eigenvalues of $\bar{\Sigma}^{-1}\Sigma_n$.

At $\bar{\Sigma} \in \mathcal{M}$, a scalar product can be defined in the associated *tangent space* $\mathcal{T}_{\bar{\Sigma}}\mathcal{M}$. This tangent space is Euclidean and locally homomorphic to the manifold. So, Riemannian distance computations in the manifold can be approximated by Euclidean distance computations in the tangent space [6].

The *logarithmic map* projects locally a $\Sigma_n \in \mathcal{M}$ onto the tangent space $\mathcal{T}_{\bar{\Sigma}}\mathcal{M}$ of $\bar{\Sigma}$ by:

$$\hat{\Sigma}_n = \mathrm{Log}_{\bar{\Sigma}}(\Sigma_n) = \bar{\Sigma}^{\frac{1}{2}}\mathrm{logm}\left(\bar{\Sigma}^{-\frac{1}{2}}\Sigma_n\bar{\Sigma}^{-\frac{1}{2}}\right)\bar{\Sigma}^{\frac{1}{2}} \tag{7}$$

where $\mathrm{logm}(\cdot)$ denotes the logarithm of a matrix [10]. The logarithm of a diagonalizable matrix $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$ is defined as $\mathrm{logm}(\mathbf{A}) = \mathbf{V}\mathbf{D}'\mathbf{V}^{-1}$, where $\mathbf{D}'$ is a diagonal matrix with elements $\mathbf{D}'_{i,i} = \log(\mathbf{D}_{i,i})$.

The *exponential map* projects an element $\hat{\Sigma}_n$ on the tangent space $\mathcal{T}_{\bar{\Sigma}}\mathcal{M}$ back to the manifold $\mathcal{M}$ by:

$$\Sigma_n = \mathrm{Exp}_{\bar{\Sigma}}(\hat{\Sigma}_n) = \bar{\Sigma}^{\frac{1}{2}}\mathrm{expm}\left(\bar{\Sigma}^{-\frac{1}{2}}\Sigma_n\bar{\Sigma}^{-\frac{1}{2}}\right)\bar{\Sigma}^{\frac{1}{2}} \tag{8}$$

where $\mathrm{expm}(\cdot)$ denotes the exponential of a matrix [10]. The exponential of a diagonalizable matrix $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$ is defined as $\mathrm{expm}(\mathbf{A}) = \mathbf{V}\mathbf{D}'\mathbf{V}^{-1}$, where $\mathbf{D}'$ is a diagonal matrix with elements $\mathbf{D}'_{i,i} = \exp(\mathbf{D}_{i,i})$.

Fig. 2 illustrates a Riemannian manifold $\mathcal{M}$, the tangent space $\mathcal{T}_{\bar{\Sigma}}\mathcal{M}$ at $\bar{\Sigma}$, the geodesic between $\bar{\Sigma}$ and Σ_n, and the corresponding logarithmic and exponential maps.

The Riemannian distance $\delta(\bar{\Sigma}, \Sigma_n)$ between two covariance matrices $\bar{\Sigma}$ and Σ_n on the manifold $\mathcal{M}$ can also be computed by a Euclidean distance in the tangent space

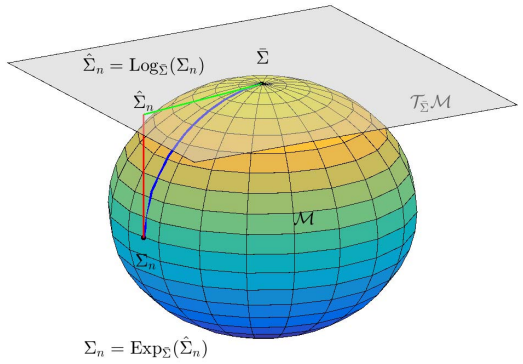


Fig. 2. Illustration of a manifold $\mathcal{M}$ and the corresponding local tangent space $\mathcal{T}_{\bar{\Sigma}}\mathcal{M}$ at $\bar{\Sigma}$. $\mathrm{Log}_{\bar{\Sigma}}(\Sigma_n)$ projects the matrix Σ_n on the manifold into the matrix $\hat{\Sigma}_n$ in the tangent space of $\bar{\Sigma}$. $\mathrm{Exp}_{\bar{\Sigma}}(\hat{\Sigma}_n)$ projects $\hat{\Sigma}_n$ in the tangent space of $\bar{\Sigma}$ into Σ_n on the manifold. The blue curve represents the geodesic between $\bar{\Sigma}$ and Σ_n on the manifold.

Algorithm 2 The Gradient Descent Algorithm for Computing the RG (Intrinsic) Mean [19]

Input: $\Sigma_n \in \mathbb{R}^{R\times R}$, $n = 1, \ldots, N$;
$\epsilon > 0$.
Output: The RG (intrinsic) mean $\bar{\Sigma} \in \mathbb{R}^{R\times R}$.
Initialize $\bar{\Sigma}_0 = \mathbf{0} \in \mathbb{R}^{R\times R}$, the zero matrix;
Initialize $\bar{\Sigma} = I \in \mathbb{R}^{R\times R}$, the identify matrix;
repeat
  $\bar{\Sigma}_0 = \bar{\Sigma}$;
  $\hat{\Sigma} = \frac{1}{N}\sum_{n=1}^{N} \mathrm{Log}_{\bar{\Sigma}_0}(\Sigma_n)$;
  $\bar{\Sigma} = \mathrm{Exp}_{\bar{\Sigma}_0}(\hat{\Sigma})$.
until $\left\|\bar{\Sigma} - \bar{\Sigma}_0\right\| < \epsilon$;
Return $\bar{\Sigma}$

around $\bar{\Sigma}$, i.e. [5],

$$\delta(\bar{\Sigma}, \Sigma_n) = \left\|\mathrm{Log}_{\bar{\Sigma}}(\Sigma_n)\right\|_{\bar{\Sigma}} = \left\|\mathrm{upper}\left(\bar{\Sigma}^{-\frac{1}{2}}\hat{\Sigma}_n\bar{\Sigma}^{-\frac{1}{2}}\right)\right\|_2 = \left\|\mathrm{upper}\left(\mathrm{logm}\left(\bar{\Sigma}^{-\frac{1}{2}}\Sigma_n\bar{\Sigma}^{-\frac{1}{2}}\right)\right)\right\|_2 \quad (9)$$

where the upper$(\cdot)$ operator keeps the upper triangular part of a symmetric matrix and vectorizes it by applying weight 1 for the diagonal elements and weight $\sqrt{2}$ for the out-of-diagonal elements [45].

The *RG mean* [42], or the *intrinsic mean* [19], of N covariance matrices is defined as the matrix minimizing the sum of the squared Riemannian distances, i.e.,

$$\bar{\Sigma} \equiv \mathfrak{G}(\Sigma_1, \ldots, \Sigma_N) = \arg\min_{\Sigma} \sum_{n=1}^{N} \delta^2(\Sigma, \Sigma_n) \quad (10)$$

There is no closed-form expression for the RG mean, but an iterative gradient descent algorithm (see Algorithm 2 [19]) can be used to find the solution. Note that Algorithm 2 makes heavy use of the logarithmic and exponential maps. In this paper we used the implementation in the Matlab Covariance Toolbox.[2]

[2]https://github.com/alexandrebarachant/covariancetoolbox.

B. Tangent Space Features for BCI Regression Problems

To use the tangent space features for BCI regression problems, we first spatially filter each $\mathbf{X}_n$ to obtain $\mathbf{X}'_n$ in (5), and then estimate its spatial covariance matrix $\Sigma_n \in \mathbb{R}^{KF\times KF}$ (note that each row of $\mathbf{X}'_n$ has zero mean):

$$\Sigma_n = \frac{1}{S}\mathbf{X}'_n\mathbf{X}'^T_n, \quad n = 1, \ldots, N \quad (11)$$

Next, we compute the Riemannian mean $\bar{\Sigma}$ of all Σ_n by Algorithm 2, and take the $KF(KF+1)/2$ upper triangular part of $\mathrm{logm}\left(\bar{\Sigma}^{-\frac{1}{2}}\Sigma_n\bar{\Sigma}^{-\frac{1}{2}}\right)$ as our features. Note that we need to assign weight 1 to the diagonal elements of $\mathrm{logm}\left(\bar{\Sigma}^{-\frac{1}{2}}\Sigma_n\bar{\Sigma}^{-\frac{1}{2}}\right)$ and weight $\sqrt{2}$ to the out-of-diagonal elements so that their Euclidean norm is equal to the Riemannian distance between $\bar{\Sigma}$ and Σ_k. The weights do not have an effect when regression methods like LASSO are used, but are very important for distance based regression methods like kNN regression.

The complete tangent space feature extraction procedure for BCI regression problems is summarized in Algorithm 3.

Algorithm 3 The Riemannian Tangent Space Feature Extraction Procedure for BCI Regression Problems

Input: Spatially filtered EEG trial $\mathbf{X}'_n \in \mathbb{R}^{KF\times S}$, $n = 1, \ldots, N$.
Output: $KF(KF+1)/2$ tangent space features for each trial.
Compute Σ_n by (11);
Compute $\bar{\Sigma}$ by Algorithm 2;
Construct the $KF(KF+1)/2$ tangent space features for $\mathbf{X}'_n$ from $\mathrm{logm}\left(\bar{\Sigma}^{-\frac{1}{2}}\Sigma_n\bar{\Sigma}^{-\frac{1}{2}}\right)$.

IV. Experiments and the Performance Evaluation Process

This section introduces a PVT experiment that was used to evaluate the performances of the proposed tangent space feature extraction method, and the corresponding RT and EEG data preprocessing procedures.

A. Experiment Setup

Seventeen university students (13 males; average age 22.4, standard deviation 1.6) from National Chiao Tung University (NCTU) in Taiwan volunteered to support the data-collection efforts over a 5-month period to study EEG correlates of attention and performance changes under specific conditions of real-world fatigue [26], as determined by the percent effectiveness score of Readiband [43]. The Institutional Review Board of NCTU approved the experimental protocol.

The customer-designed daily sampling system consists of a smartphone, actigraph, sleep diary, subjective scales of fatigue and stress, and software for recording, storing, transmitting, and analyzing data acquired from individuals in their natural environments on a daily basis. Each participant was provided a wrist-worn actigraph (Fatigue Science Readiband,

TABLE I
NUMBER OF TRIALS AND MEAN RTS FOR THE 17 SUBJECTS

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Number of trials	809	813	839	685	843	465	833	610	769	813	743	828	803	794	528	823	553
Mean RT (s)	0.47	0.51	0.47	0.98	0.49	0.42	0.61	0.44	0.81	0.46	1.03	0.45	0.57	0.59	0.67	0.49	2.14

Vancouver, BC), and was instructed to complete several subjective report scales and enter the percent effectiveness score from the actigraph approximately 30-60 minutes upon awakening each morning and to be available for experiment testing approximately once every 1-3 weeks over a 5-month period for a total of nine repeated sessions. Data recorded by the daily sampling system included electronically-adapted visual analog scales of fatigue and stress, the Karolinska Sleepiness Scale [1], and the Pittsburgh Sleep Diary [38]. The daily sampling data were automatically uploaded from the smartphone to a designated secure server at NCTU on a daily basis. In this way we could track and identify periods when the participants were currently exhibiting low, normal, or high levels of fatigue based on the percent effectiveness score values ($>90\%$, $70-90\%$, $<70\%$, respectively). The goal was to examine the participants during experiment sessions three times within each of the three fatigue levels. Most participants finished all nine sessions.

When the participants reported to the laboratory, we measured their fatigue level on site again right before the experiment to make sure it was close to the fatigue state reported via the smartphone. Upon completion of the related questionnaires and the informed consent form, subjects performed a PVT, a dynamic attention-shifting task, a lane-keeping task, and selected surveys preceding each condition. EEG data were recorded at 1000 Hz using a 64-channel NeuroScan Quik-Cap system (62 EEG channels and 1 electrocardiogram channel). The ground was between FPZ and FZ, and the reference channels were A1 and A2 at the mastoids.

In this paper we focus on the PVT [15], which is a sustained-attention task that uses RT to measure the speed with which a subject responds to a visual stimulus. It is widely used, particularly by NASA, for its ease of scoring, simple metrics, convergent validity, and free of learning effects. In our experiment, the PVT was presented on a smartphone with each trial initiated as an empty solid white circle centered on the touchscreen that began to fill in red displayed as a clockwise sweeping motion like the hand of a clock. The sweeping motion was programmed to turn solid red in one second or terminate upon a response by the participants, which required them to tap the touchscreen with the thumb of their dominant hand. The RT was computed as the elapsed time between the appearance of the empty solid white circle and the participant's response. Following completion of each trial, the circle went back to solid white until the next trial. Inter-trial intervals consisted of random intervals between 2-10 seconds.

143 sessions of PVT data were collected from the 17 subjects, and each session lasted 10 minutes. Our goal is to predict the RT using a short EEG trial immediately before it.

B. Performance Evaluation Process

The following procedure was used to evaluate the performances of different feature extraction methods:

1) *RT data preprocessing to remove outliers.*
The number of trials and the mean RTs for the 17 subjects are shown in Table I. Subject 17 may have data recording issues, because many of his RTs were longer than 5 seconds, which are highly unlikely in practice, and his mean RT was more than two times larger than the largest mean RT from other subjects. So we excluded him from consideration in this paper, and only used Subjects 1-16.
The RTs were very noisy, and there were obvious outliers. It is very important to suppress the outliers and noise so that the performances of different algorithms can be more accurately compared. We employed the following 2-step procedure for RT data preprocessing:
   a) *Outlier removal*, which aimed to remove abnormally large RTs. First, a threshold $\theta = m_y + 3\sigma_y$ was computed for each subject, where m_y is the mean RT from all sessions of that subject, and σ_y is the corresponding standard deviation. Then, all RTs larger than θ were removed. Note that the threshold was different for different subjects.
   b) *Moving average smoothing*, which replaced each RT by the average RT during a 60 seconds moving window centered at the onset of the corresponding PVT to suppress the noise.
2) *EEG data preprocessing to remove or suppress artifacts and noise.*
Generally raw EEG data recorded from the scalp contain many artifacts (e.g., head motion, blinks, eye movements, etc.) and noise (e.g., power-line noise, noise caused by changes in electrode impedances, etc.) [11], [46], so it is very important to remove or suppress them to increase the signal-to-noise ratio before a machine learning algorithm is applied. This paper used the standardized early-stage EEG processing pipeline (PREP) [11], which consists of three steps: a) remove line-noise, b) determine and remove a robust reference signal, and, c) interpolate the bad channels (channels with a low recording signal-to-noise ratio).
The preprocessed EEG signals coming out of PREP were downsampled to 250 Hz. They were then epoched to 5-second trials according to the onset of the PVTs: if a PVT started at t, then the 62-channel EEG trial in $[t-5, t]$ seconds was used to predict the RT, i.e., $\mathbf{X}_n \in \mathbb{R}^{62\times 1250}$. Each trial was then individually filtered by a $[1, 20]$ Hz finite impulse response band-pass

filter to make each channel zero-mean and to remove non-relevant high frequency components.

3) *5-fold cross-validation to compute the regression performance for each combination of feature set and regression method.*

We first randomly partitioned the trials into five folds; then, used four folds for supervised spatial filtering and regression model training, and the remaining fold for testing. We repeated this five times so that every fold was used in testing. Finally we computed the regression performances in terms of root mean square error (RMSE) and correlation coefficient (CC).

We extracted the following three different feature sets for each preprocessed EEG trial:

- *Feature Set 1 (FS1): Theta and Alpha powerband features from the band-pass filtered EEG trials.* We computed the average power spectral density in the Theta band (4-8 Hz) and Alpha band (8-13 Hz) for each channel using Welch's method [50], and converted these $62 \times 2 = 124$ band powers to dBs as our features.
- *Feature Set 2 (FS2): Theta and Alpha powerband features from EEG trials filtered by Algorithm 1.* This procedure was almost identical to the above one, except that the band-pass filtered EEG trials were also spatially filtered by Algorithm 1 before the powerband features were computed. We used 3 fuzzy sets for the RTs, and 10 spatial filters for each fuzzy class, so that the spatially filtered EEG trials had dimensionality 30×1250, and FS2 had 60 dimensions.
- *Feature Set 3 (FS3): Riemannian tangent space features from EEG trials filtered by Algorithm 1.* That is, we first band-pass filtered the raw EEG signals, then spatially filtered them by Algorithm 1 ($K = 10$ and $F = 3$), and further applied Algorithm 3 to extract the tangent space features, which had $30 \times 31/2 = 465$ dimensions.

Two regression methods were used on each feature set: LASSO [44], and kNN regression [2].

For labeled training data $\{\mathbf{x}_n, y_n\}_{n=1,\ldots,N}$, LASSO solves the following minimization problem to find a sparse linear regression model:

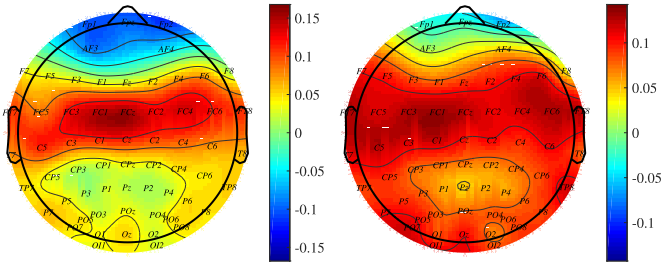
$$\min_{\beta_0,\boldsymbol{\beta}} \left[\frac{1}{2N} \sum_{n=1}^{N} \left(y_n - \beta_0 - \boldsymbol{\beta}^T \mathbf{x}_n \right)^2 + \lambda \left\| \boldsymbol{\beta} \right\|_1 \right] \quad (12)$$

where $\lambda > 0$ is an adjustable parameter, which was optimized by an inner 5-fold cross-validation on the training dataset in this paper. Once β_0 and $\boldsymbol{\beta}$ are identified, the final LASSO regression model is:

$$\hat{y}_n = \beta_0 + \boldsymbol{\beta}^T \mathbf{x}_n \quad (13)$$

We used $k = 5$ in kNN. Once the five nearest neighbors $\{\mathbf{x}_i, y_i\}_{i=1,\ldots,5}$ to the new trial $\mathbf{x}_n$ are identified, the regression output is computed as a weighted average:

$$\hat{y}_n = \frac{\sum_{i=1}^{5} w_i y_i}{\sum_{i=1}^{5} w_i} \quad (14)$$



(a) (b)

Fig. 3. Topoplot of the average CC between the RT and the powerband features from FS1 at different channel locations. (a) theta; (b) alpha.

where the weights are the inverses of the feature distances:

$$w_i = \frac{1}{\left\| \mathbf{x}_n - \mathbf{x}_i \right\|_2} \quad (15)$$

4) *Repeat Step 3 10 times and compute the average regression performance.*

V. Experimental Results

This section compares the informativeness of the features in FS1, FS2 and FS3, and presents the regression performances.

A. Informativeness of the Features

Before studying the regression performance, it is important to check if the extracted features in FS1, FS2 and FS3 are indeed meaningful.

In this first study, we computed the CC between the RT and powerband features in FS1 at different channel locations for each of the 16 subjects, and then averaged them. The corresponding topoplot is shown in Fig. 3. Both theta and alpha band powers show higher correlation at the central and central-frontal regions of the brain; however, generally the CC is small. This indicates that FS1 features are not very informative.

In the second study, we picked a typical subject, partitioned his data randomly into 50% training and 50% testing, and extracted the powerband features FS1. We then designed the spatial filters using Algorithm 1 on the training data, and extracted the corresponding powerband features FS2, and the Riemannian tangent space features FS3 using Algorithm 3. For each feature set, we identified the top three features that had the maximum CCs with the RT using the training data, and also computed the corresponding CCs for the testing data. The results are shown in Fig. 4, where in each panel the data on the left of the black dotted line were used for training, and the right for testing. The top thick curve is the RT, and the bottom three curves are the maximally correlated features identified from the training data. The training and testing CCs are shown on the left and right of the corresponding feature, respectively. For FS1, we also show the corresponding channel labels and powerband names. For FS2, we only show the powerband names of the top three features, as a channel here does not have a specific label (each channel in FS2 is a weighted combination of all 62 physical electrodes).

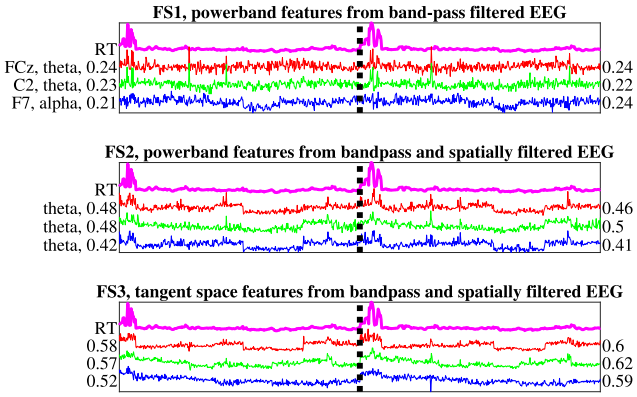


Fig. 4. Features from different feature extraction methods, and the corresponding training and testing CCs with the RT.

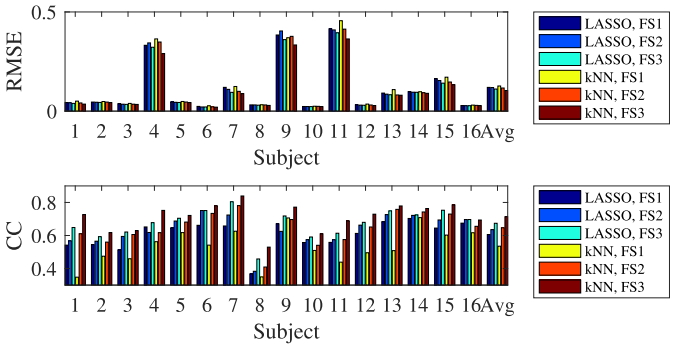


Fig. 5. RMSEs and CCs of the six approaches on the 16 subjects.

Fig. 4 shows that FS2 gave much smoother features than FS1, and also achieved much larger CCs to the RT, both in training and testing, suggesting that spatial filtering by Algorithm 1 can indeed increase the signal quality. FS3 further achieved larger training and testing CCs to the RT than FS2, suggesting that the tangent space features are more informative than the powerband features.

B. Estimation Performance Comparison

The RMSEs and CCs of LASSO and kNN using three different feature sets are shown in Fig. 5 for the 16 subjects. Recall that for each subject the feature extraction methods were run 10 times, each with randomly partitioned training and testing data, and the average regression performances are shown here. The average RMSEs and CCs across all subjects are also shown in the last group of each panel.

Fig. 5 shows that regardless of which regression method was used, generally FS2 resulted in smaller RMSEs and larger CCs than FS1, suggesting that the spatial filtering approach can indeed improve the regression performance. Fig. 5 also shows that FS3 further achieved better RMSEs and CCs than FS2, suggesting that the tangent space features were more effective than the powerband features. Finally, LASSO had better performance than kNN on FS1, but kNN became better on FS2 and FS3. The RMSEs for Subjects 4, 9 and 11 in Fig. 5 are much larger than others, because, as shown in Table I, these three subjects have much larger RTs than others.

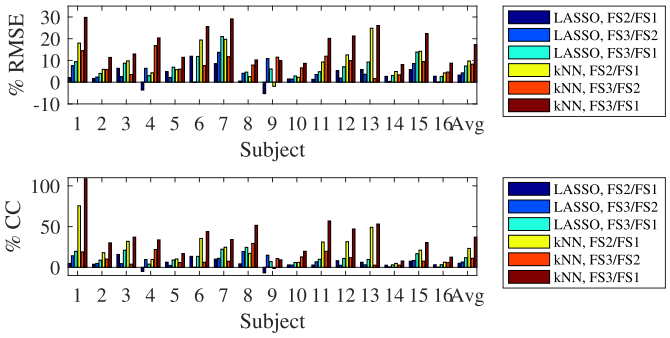


Fig. 6. Pairwise percentage performance improvement of the algorithms on the 16 subjects.

TABLE II
p-VALUES OF TWO-WAY ANOVA TESTS FOR {FS1, FS2, FS3}

	LASSO		kNN	
	RMSE	CC	RMSE	CC
p for raw values	.8183	**.0000**	.2742	**.0000**
p for ratios	**.0000**	**.0000**	**.0000**	**.0000**

To illustrate the performance differences among the three feature extraction methods from another viewpoint, Fig. 6 shows the corresponding percentage performance improvements of LASSO and kNN using the three feature sets, where the legend "LASSO, FS2/FS1" means the percentage performance improvement of LASSO on FS2 over LASSO on FS1, and other legends should be understood in a similar manner. For LASSO, on average FS3 had 4.30% smaller RMSE than FS2, and 6.59% larger CC. For kNN, on average FS3 had 8.30% smaller RMSE than FS2, and 11.13% larger CC. These results again demonstrated that the tangent space features are more effective than the traditional powerband features.

We also performed a two-way Analysis of Variance (ANOVA) for different regression algorithms to check if the raw RMSE and CC differences among the three feature sets (FS1, FS2, and FS3) were statistically significant, by setting the subjects as a random effect. The results are shown in Table II as "p for raw values." Study results showed that there were statistically significant differences (at 5% level) in raw CCs among different feature sets for both LASSO and kNN, but not for raw RMSEs.

However, because the RTs from different subjects had significantly different magnitudes, an ANOVA on the raw RMSEs and CCs may be unfair for those subjects with small RTs. So, we also performed a two-way ANOVA for different algorithms and feature sets on the ratios. For example, to compute the RMSE ratios for LASSO, we replaced all RMSEs for FS1 by 1, the RMSEs for FS2 by the ratios of the corresponding RMSEs from FS2 over those from FS1, and the RMSEs for FS3 by the ratios of the corresponding RMSEs from FS3 over those from FS1. In this way the RMSEs were normalized, and hence different subjects were treated equally. The corresponding ANOVA test results are shown in Table II as "p for ratios." Observe that there were statistically significant differences (at 5% level) in both RMSE ratios and CC ratios among different feature sets for both LASSO and kNN.

TABLE III
p-VALUES OF NON-PARAMETRIC MULTIPLE COMPARISON ON THE RAW VALUES FOR {FS1, FS2, FS3}

	LASSO				kNN			
	RMSE		CC		RMSE		CC	
	FS1	FS2	FS1	FS2	FS1	FS2	FS1	FS2
FS2	.2143		**.0001**		.0852		**.0000**	
FS3	.1319	.2702	**.0000**	**.0001**	**.0034**	.0711	**.0000**	**.0000**

TABLE IV
p-VALUES OF NON-PARAMETRIC MULTIPLE COMPARISON ON THE RATIOS FOR {FS1, FS2, FS3}

	LASSO				kNN			
	RMSE		CC		RMSE		CC	
	FS1	FS2	FS1	FS2	FS1	FS2	FS1	FS2
FS2	**.0000**		**.0000**		**.0000**		**.0000**	
FS3	**.0000**	**.0000**	**.0000**	**.0000**	**.0000**	**.0000**	**.0000**	**.0000**

Then, non-parametric multiple comparison tests based on Dunn's procedure [17], [18] were used to determine if the difference between any pair of algorithms was statistically significant, with a p-value correction using the False Discovery Rate method [9]. The p-values for the raw values are shown in Table III, and the p-values for the ratios are shown in Table IV, where the statistically significant ones are marked in bold. Table III shows that the raw RMSE difference between FS3 and FS1 was statistically significant when kNN was used. Furthermore, the raw CC differences between all pairs of feature sets were statistically significant. Table IV shows that the ratio differences between all pairs of feature sets were statistically significant, for both LASSO and kNN.

VI. DISCUSSIONS

This section provides parameter sensitivity analysis and additional discussions.

A. Parameter Sensitivity Analysis

Tangent space feature extraction relies on the spatial filter in Algorithm 1, which has two adjustable parameters: K, the number of fuzzy classes for the RTs, and F, the number of spatial filters for each fuzzy class. The filtering performance is robust to K but changes noticeably when F changes [55]. As a result, the performance of the tangent space features also varies as F changes. In this subsection we study the sensitivity of the regression performance to F.

The regression performances for $F = \{5, 10, 15, 20\}$ (K was fixed to be 3) are shown in Fig. 7. Algorithms 1 and 3 were repeated five times, each time with a random partition of training and testing data, and the average regression results are shown. Note that F cannot be too large because of three constraints: 1) F cannot exceed the number of channels (C) in the original EEG data, because $\bar{\Sigma}_k \bar{\Sigma}^{-1} \in \mathbb{R}^{C\times C}$ in (3) has at most C eigenvectors; 2) the tangent space features have dimensionality $KF(KF+1)/2$, which increases rapidly with F; so, a large F can easily result in over-fitting; and, 3) there may be numerical difficulties in computing the RG mean when F is large, e.g., for Subjects 5, 8 and 15 in Fig. 7 when $F = 20$.

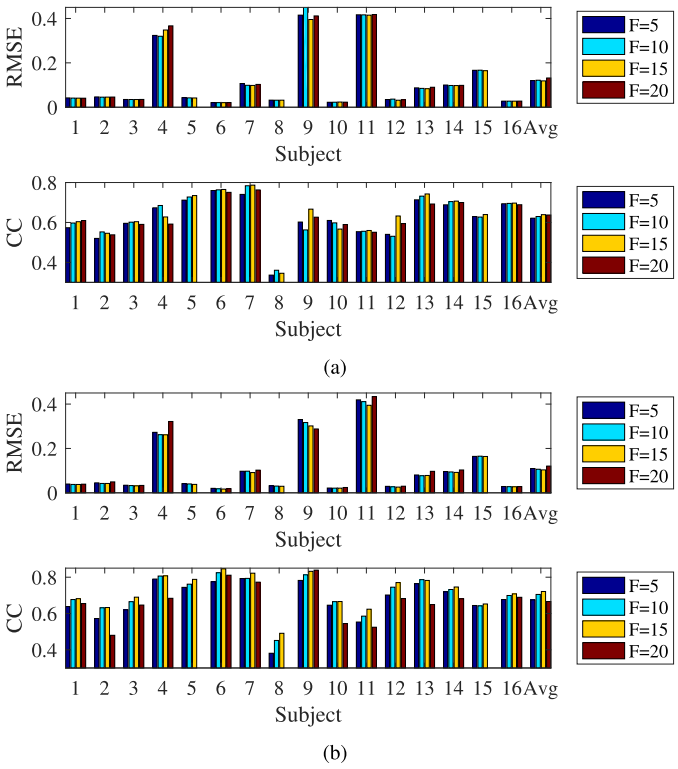


Fig. 7. RMSEs and CCs of (a) LASSO and (b) kNN with respect to F, the number of spatial filters for each fuzzy class in Algorithm 1.

Fig. 7 shows that the regression performance increased when F increased from 5 to 15, but decreased when F further increased to 20. For the PVT experiment, $F \in [10, 15]$ seemed to achieve a good compromise between performance and computational cost.

Additionally, in the previous subsection we used 5-second EEG trials to estimate the corresponding RT, and it is also interesting to study how the estimation performance changes with different trial lengths. The results are shown in Fig. 8 for trial lengths of $\{1, 3, 5, 7, 9\}$ seconds. In general, as trial length increased, the estimation performance improved. However, a longer trial means heavier computational cost and larger delay in estimation. Furthermore, a trial cannot be arbitrary long, as then it cannot capture the up-to-date RT. These effects should be taken into consideration when choosing the right trial length.

B. Regression Performance Versus the Number of Features

Recall from Section IV-B that FS1 has 124 features, FS2 has 60 features, and FS3 has 465 features, i.e., FS3 has much more features than FS1 and FS2. So, FS3's superior performance may be due to its increased number of features. In this subsection we investigate the relationship between the regression performance and the number of useful features.

Because LASSO automatically selects the most useful features, whereas kNN always uses all the features, in this study we focus only on LASSO. For each subject and each feature set, we used all data in LASSO training, and recorded the number of selected features, as well as the corresponding

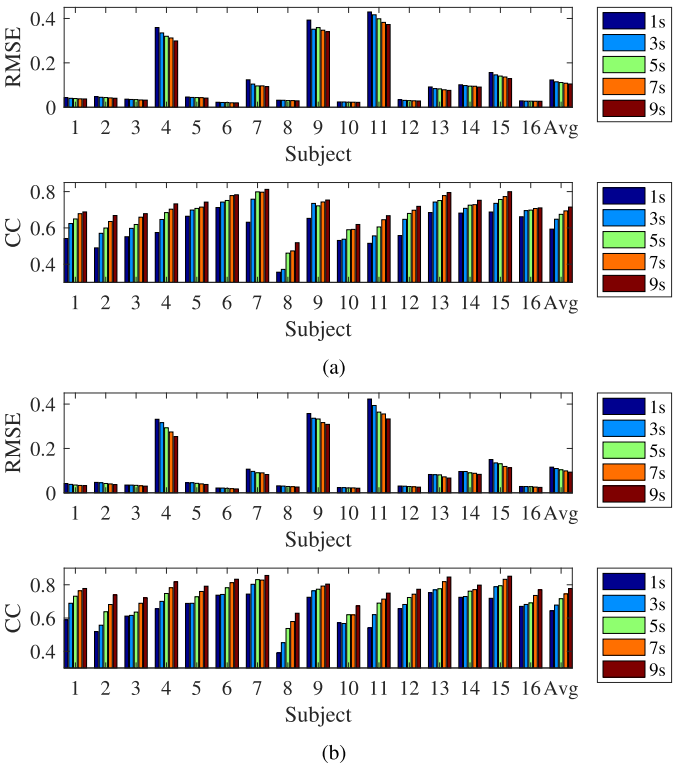


Fig. 8. RMSEs and CCs of (a) LASSO and (b) kNN with respect to the trial length.

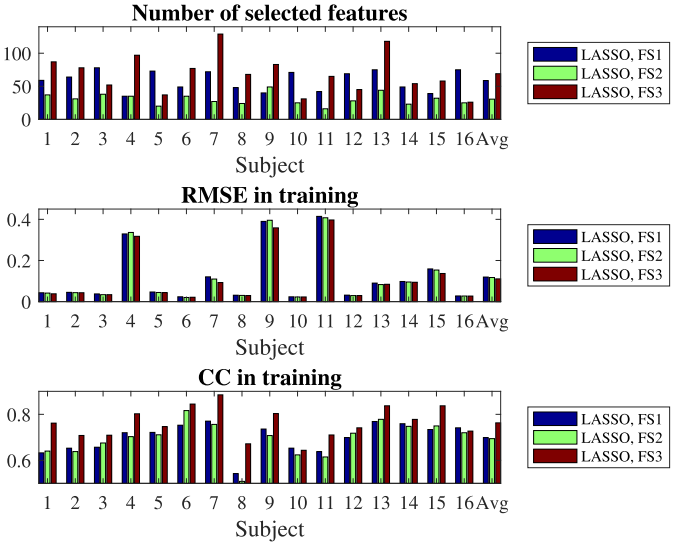


Fig. 9. The nubmer of features selected by LASSO, and the corresponding training RMSEs and CCs.

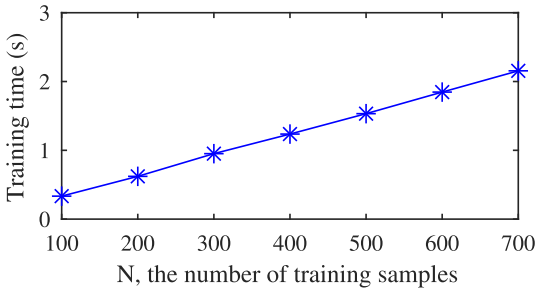


Fig. 10. The training time of our feature extraction method w.r.t. N.

training RMSEs and CCs. The results are shown in Fig. 9. On average LASSO selected 58.6 features from `FS1`, 30.6 features from `FS2`, and 69.1 features from `FS3`. Although the selected `FS2` subset was only about half the size of the selected `FS1` subset, they resulted in similar overall training RMSEs and CCs. Connecting this observation with that in the previous subsection, i.e., `FS2` had much better testing RMSEs and CCs than `FS1`, we can conclude that the CSPR-OVR spatial filter can aggregate the most useful information into just a small number of features, which reduces overfitting and improves the generalization performance. Fig. 9 also shows that the selected `FS3` subset was slightly larger than the selected `FS1` subset, but the `FS3` subset resulted in much better training performance, and also much better testing performance, as presented in the previous subsection. These observations together suggest that the Riemannian geometry approach can indeed extract some novel informative features, which improve both the training and the testing performances.

C. Computational Cost

The training of our feature extraction method (`FS3`) consists of three steps: 1) design the CSPR-OVR filter by Algorithm 1; 2) compute the RG mean $\bar{\Sigma}$ by Algorithm 2; and, 3) map the spatially filtered EEG trials to the Riemannian tangent space by Algorithm 3. Once the training is done, feature extraction for a testing trial can be performed very efficiently: a matrix multiplication (5) is first used to spatially filter it, and then another matrix multiplication (11) is used to compute its spatial covariance matrix Σ_n; finally, compute $\log\mathrm{m}\left(\bar{\Sigma}^{-\frac{1}{2}}\Sigma_n\bar{\Sigma}^{-\frac{1}{2}}\right)$ and take its upper triangular part as the features. Note that $\bar{\Sigma}$ has been obtained in training, so $\bar{\Sigma}^{-\frac{1}{2}}$ can be pre-computed, and hence $\bar{\Sigma}^{-\frac{1}{2}}\Sigma_n\bar{\Sigma}^{-\frac{1}{2}}$ is also a simple matrix multiplication. So, in this subsection we focus on the training computational cost only.

Let N be the number of training samples. Then, the actual training time increased linearly with N, as shown in Fig. 10. The platform was a Dell XPS15 laptop (Intel i7-6700HQ CPU @2.60GHz, 16 GB memory) running Windows 10 Pro 64-bit and Matlab 2016b. A least squares curve fit shows that the training time is $0.0261 + 0.0030N$ seconds, which should not be a problem for a practical N.

D. RT Versus Fatigue State

We also studied the relationship between the RT and the fatigue state. Our conjecture is that as the fatigue level goes up, the RT should be larger. Boxplots of the RT in different sessions for two typical subjects are shown in Fig. 11, where "L", "N" and "H" mean low, normal, and high fatigue, respectively. Fig. 11 shows that the mean RT of a high fatigue sessions is generally larger than that of a low or normal fatigue session, and the former also has more extreme values and a larger variance. The difference between a low fatigue session and a normal fatigue session is not obvious. These observations suggest that although the fatigue state

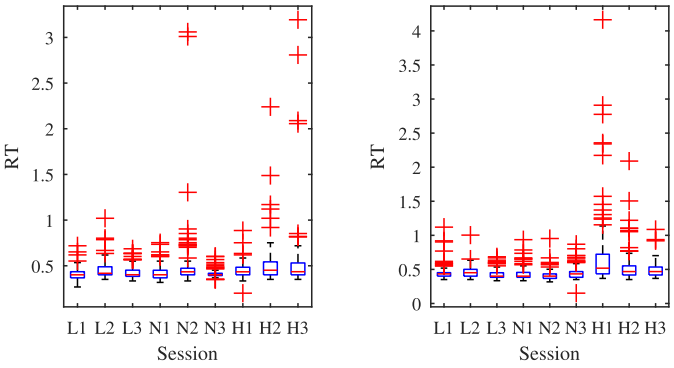


Fig. 11. Boxplots of the RT in different fatigue states for two typical subjects.

contains some useful information, it may be too coarse for accurate RT prediction. That's why it was not used in this paper.

VII. Conclusions and Future Research

In this paper, we have proposed a new feature extraction approach for EEG-based BCI regression problems: a spatial filter is first used to increase the EEG trial signal quality and also to reduce the dimensionality of the covariance matrix, and then Riemannian tangent space features are extracted. We validated the performance of the proposed approach in RT estimation from EEG signals measured in a large-scale sustained-attention PVT experiment, and showed that compared with the traditional powerband features, the tangent space features can reduce the estimation RMSE by 4.30-8.30%, and increase the estimation CC by 6.59-11.13%. To our knowledge, this is the first time that RG has been used in BCI regression problems.

Our future research will focus on reducing the dimensionality of the tangent space features. As shown in Algorithm 3, the tangent space features have dimensionality $KF(KF+1)/2$, where K is the number of fuzzy classes for the RTs, and F is the number of spatial filters for each fuzzy class. So, the feature dimensionality increases quadratically with respect to both K and F, which quickly results in overwhelming computational cost, overfitting, and numerical problems. We will investigate effective dimensionality reduction approaches for the tangent space features to reduce the computational cost while maintaining or even improving the regression performance.

Acknowledgment

The views and the conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory or the U.S. Government.

References

[1] T. Åkerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," *Int. J. Neurosci.*, vol. 52, nos. 1–2, pp. 29–37, 1990.

[2] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Amer. Statist.*, vol. 46, no. 3, pp. 175–185, 1992.

[3] S. Amari, H. Nagaoka, and D. Harada, *Methods of Information Geometry*. Providence, RI, USA: AMS, 2000.

[4] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Geometric means in a novel vector space structure on symmetric positive-definite matrices," *SIAM J. Matrix Anal. Appl.*, vol. 29, no. 1, pp. 328–347, 2007.

[5] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain–computer interface classification by Riemannian geometry," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 4, pp. 920–928, Apr. 2012.

[6] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications," *Neurocomputing*, vol. 112, pp. 172–178, Jul. 2013.

[7] A. Barachant. (2014). *MEG Decoding Using Riemannian Geometry and Unsupervised Classification*, accessed on Aug. 17, 2016. [Online]. Available: http://alexandre.barachant.org/wp-content/uploads/2014/08/documentation.pdf

[8] A. Barachant and M. Congedo. (Aug. 2014). "A plug & play P300 BCI using information geometry." [Online]. Available: https://arxiv.org/abs/1409.0107

[9] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Statist. Soc. Ser. B, Methodol.*, vol. 57, no. 1, pp. 289–300, 1995.

[10] M. Berger, *A Panoramic View of Riemannian Geometry*. New York, NY, USA: Springer, 2007.

[11] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K.-M. Su, and K. A. Robbins, "The PREP pipeline: Standardized preprocessing for large-scale EEG analysis," *Frontiers Neuroinform.*, vol. 9, p. 16, Jun. 2015.

[12] T. J. Bradberry, R. J. Gentili, and J. L. Contreras-Vidal, "Reconstructing three-dimensional hand movements from noninvasive electroencephalographic signals," *J. Neurosci.*, vol. 30, no. 9, pp. 3432–3437, 2010.

[13] T. J. Bradberry, R. J. Gentili, and J. L. Contreras-Vidal, "Fast attainment of computer cursor control with noninvasively acquired brain signals," *J. Neural Eng.*, vol. 8, no. 3, p. 036010, 2011.

[14] M. Congedo, A. Barachant, and A. Andreev. (Oct. 2013). "A new generation of brain-computer interface based on Riemannian geometry." [Online]. Available: https://arxiv.org/abs/1310.8115

[15] D. F. Dinges and J. W. Powell, "Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations," *Behavior Res. Methods, Instrum., Comput.*, vol. 17, no. 6, pp. 652–655, 1985.

[16] S. P. Drummond, A. Bischoff-Grethe, D. F. Dinges, L. Ayalon, S. C. Mednick, and M. Meloy, "The neural basis of the psychomotor vigilance task," *Sleep*, vol. 28, no. 9, pp. 1059–1068, 2005.

[17] O. J. Dunn, "Multiple comparisons among means," *J. Amer. Statist. Assoc.*, vol. 56, pp. 62–64, Mar. 1961.

[18] O. J. Dunn, "Multiple comparisons using rank sums," *Technometrics*, vol. 6, no. 3, pp. 214–252, 1964.

[19] P. T. Fletcher and S. Joshi, "Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors," in *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis* (Lecture Notes in Computer Science), vol. 3117. Berlin, Germany: Springer, 2004, pp. 87–98.

[20] W. Forstner and B. Moonen, "A metric for covariance matrices," in *Geodesy—The Challenge of the 3rd Millennium*, E. W. Grafarend, F. W. Krumm and V. S. Schwarze, Eds., Berlin, Germany: Springer, 2003, pp. 299–309.

[21] J. Fruitet, D. J. McFarland, and J. R. Wolpaw, "A comparison of regression techniques for a two-dimensional sensorimotor rhythm-based brain–computer interface," *J. Neural Eng.*, vol. 7, no. 1, p. 016003, 2010.

[22] G. H. Golub and C. F. V. Loan, *Matrix Computation*, 3rd ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996.

[23] B. He, B. Baxter, B. J. Edelman, C. C. Cline, and W. W. Ye, "Noninvasive brain-computer interfaces based on sensorimotor rhythms," *Proc. IEEE*, vol. 103, no. 6, pp. 907–925, Jun. 2015.

[24] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, "Transfer learning in brain-computer interfaces," *IEEE Comput. Intell. Mag.*, vol. 11, no. 1, pp. 20–31, Feb. 2016.

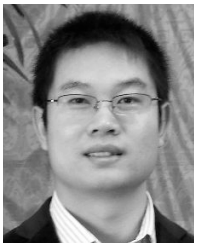
[25] E. K. Kalunga, S. Chevallier, Q. Barthélemy, K. Djouani, E. Monacelli, and Y. Hamam, "Online SSVEP-based BCI using Riemannian geometry," *Neurocomputing*, vol. 191, pp. 55–68, May 2016.

[26] S. Kerick *et al.*, "Inter- and intra-individual variations in sleep, subjective fatigue, and vigilance task performance of students in their real-world environments over extended periods," to be published.

[27] J. Lee, *Introduction to Smooth Manifolds*. New York, NY, USA: Springer, 2002.

[28] Y. Li, K. Wong, and H. de Bruin, "Electroencephalogram signals classification for sleepstate decision—A riemannian geometry approach," *IET Signal Process.*, vol. 6, no. 4, pp. 288–299, 2012.

[29] L.-D. Liao *et al.*, "Biosensor technologies for augmented brain–computer interfaces in the next decades," *Proc. IEEE*, vol. 100, no. 2, pp. 1553–1566, May 2012.
[30] C.-T. Lin, R.-C. Wu, S.-F. Liang, W.-H. Chao, Y.-J. Chen, and T.-P. Jung, "EEG-based drowsiness estimation for safety driving using independent component analysis," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 52, no. 12, pp. 2726–2738, Dec. 2005.
[31] C. T. Lin *et al.*, "Development of wireless brain computer interface with embedded multitask scheduling and its application on real-time driver's drowsiness detection and warning," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 5, pp. 1582–1591, May 2008.
[32] C.-T. Lin *et al.*, "Adaptive EEG-based alertness estimation system by using ICA-based fuzzy neural networks," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 11, pp. 2469–2476, Nov. 2006.
[33] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces," *Proc. IEEE*, vol. 103, no. 6, pp. 871–890, Jun. 2015.
[34] S. Makeig, C. Kothe, T. Mullen, N. Bigdely-Shamlo, Z. Zhang, and K. Kreutz-Delgado, "Evolving signal processing for brain–computer interfaces," *Proc. IEEE*, vol. 100, Special Centennial Issue, pp. 1567–1584, May 2012.
[35] D. J. McFarland, A. T. Lefkowicz, and J. R. Wolpaw, "Design and operation of an EEG-based brain-computer interface with digital signal processing technology," *Behavior Res. Methods, Instrum., Comput.*, vol. 29, no. 3, pp. 337–345, 1997.
[36] D. J. McFarland, W. A. Sarnacki, and J. R. Wolpaw, "Electroencephalographic (EEG) control of three-dimensional movement," *J. Neural Eng.*, vol. 7, no. 3, p. 036007, 2010.
[37] M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 3, pp. 735–747, 2005.
[38] T. Monk *et al.*, "The Pittsburgh sleep diary," *J. Sleep Res.*, vol. 3, no. 2, pp. 111–120, 1994.
[39] X. Navarro-Sune *et al.*, "Riemannian geometry applied to detection of respiratory states from EEG signals: The basis for a brain–ventilator interface," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 5, pp. 1138–1148, May 2017.
[40] F. Parisi, F. Strino, B. Nadler, and Y. Kluger, "Ranking and combining multiple predictors without labeled data," *Proc. Nat. Acad. Sci.*, vol. 111, no. 4, pp. 1253–1258, 2014.
[41] X. Pennec, "Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements," *J. Math. Imag. Vis.*, vol. 25, no. 1, pp. 127–154, 2006.
[42] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *Int. J. Comput. Vis.*, vol. 66, no. 1, pp. 41–66, 2006.
[43] C. Russell, J. Caldwell, D. Arand, L. Myers, P. Wubbels, and H. Downs. (2015). *Validation of the Fatigue Science Readiband Actigraph and Associated Sleep/Wake Classification Algorithms*. accessed on Aug. 11, 2016. http://static1.squarespace.com/static/550af02ae4b0cf85628d981a/t/5526c99ee4b019412c323758/1428605342303/Readiband_Validation.pdf
[44] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc., B, Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.
[45] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on Riemannian manifolds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1713–1727, Oct. 2008.
[46] J. A. Urigüen and B. Garcia-Zapirain, "EEG artifact removal—State-of-the-art and guidelines," *J. Neural Eng.*, vol. 12, no. 3, p. 031001, 2015.
[47] P. Wang, J. Lu, B. Zhang, and Z. Tang, "A review on transfer learning for brain-computer interface classification," in *Proc. 5th Int. Conf. Inf. Sci. Technol. (ICIST)*, Changsha, China, Apr. 2015, pp. 315–322.
[48] N. R. Waytowich, V. J. Lawhern, A. W. Bohannon, K. R. Ball, and B. J. Lance, "Spectral transfer learning using information geometry for a user-independent brain-computer interface," *Frontiers Neurosci.*, vol. 10, p. 430, Sep. 2016.
[49] C.-S. Wei, Y.-P. Lin, Y.-T. Wang, T.-P. Jung, N. Bigdely-Shamlo, and C.-T. Lin, "Selective transfer learning for EEG-based drowsiness detection," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Hong Kong, Oct. 2015, pp. 3229–3232.
[50] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, Jun. 1967.
[51] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, "An EEG-based brain-computer interface for cursor control," *Electroencephalograph. Clin. Neurophysiol.*, vol. 78, no. 3, pp. 252–259, 1991.
[52] J. R. Wolpaw, D. J. McFarland, and T. M. Vaughan, "Brain-computer interface research at the Wadsworth center," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 222–226, Jun. 2000.
[53] J. Wolpaw and N. Birbaumer, "Brain-computer interfaces for communication and control," in *Textbook for Neural Repair and Repair and Rehabilitation*, M. Selzer, L. Cohen, F. Gage, S. Clarke, and P. Duncan, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2006, pp. 602–614.
[54] D. Wu, C.-H. Chuang, and C.-T. Lin, "Online driver's drowsiness estimation using domain adaptation with model fusion," in *Proc. Int. Conf. Affective Comput. Intell. Interact.*, Xi'an, China, Sep. 2015, pp. 904–910.
[55] D. Wu, J.-T. King, C.-H. Chuang, C.-T. Lin, and T.-P. Jung, "Spatial filtering for EEG-based regression problems in brain-computer interface (BCI)," *IEEE Trans. Fuzzy Syst.*, to be published. [Online]. Available: https://arxiv.org/abs/1702.02914
[56] D. Wu, V. J. Lawhern, S. Gordon, B. J. Lance, and C.-T. Lin, "Driver drowsiness estimation from EEG signals using online weighted adaptation regularization for regression (OwARR)," *IEEE Trans. Fuzzy Syst.*, to be published.
[57] D. Wu, V. J. Lawhern, S. Gordon, B. J. Lance, and C.-T. Lin, "Offline EEG-based driver drowsiness estimation using enhanced batch-mode active learning (EBMAL) for regression," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Budapest, Hungary, Oct. 2016, pp. 730–736.
[58] D. Wu, V. J. Lawhern, S. Gordon, B. J. Lance, and C.-T. Lin, "Spectral meta-learner for regression (SMLR) model aggregation: Towards calibrationless brain-computer interface (BCI)," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Budapest, Hungary, Oct. 2016, pp. 743–749.
[59] F. Yger, F. Lotte, and M. Sugiyama, "Averaging covariance matrices for EEG signal classification based on the CSP: An empirical study," in *Proc. 23rd Eur. Signal Process. Conf. (EUSIPCO)*, Nice, France, Aug. 2015, pp. 2721–2725.
[60] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965.



Dongrui Wu (S'05–M'09–SM'14) received the B.E. degree in automatic control from the University of Science and Technology of China, Hefei, China, in 2003, the M.Eng. degree in electrical engineering from the National University of Singapore, Singapore, in 2005, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 2009.

He was a Lead Research Engineer with GE Global Research from 2010 to 2015. He is currently a Chief Scientist with DataNova, NY, USA. He has authored or co-authored over 90 publications, including the book *Perceptual Computing* (Wiley-IEEE, 2010). His current research interests include affective computing, brain–computer interface, computational intelligence, and machine learning.

Dr. Wu received the IEEE International Conference on Fuzzy Systems Best Student Paper Award in 2005, the IEEE Computational Intelligence Society Outstanding Ph.D. Dissertation Award in 2012, the IEEE TRANSACTIONS ON FUZZY SYSTEMS Outstanding Paper Award in 2014, and the North American Fuzzy Information Processing Society Early Career Award in 2014. He was a finalist of the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING Most Influential Paper Award in 2015 and the IEEE Brain Initiative Best Paper Award in 2016. He is an Associate Editor of the IEEE TRANSACTIONS ON FUZZY SYSTEMS, the IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS, and the *IEEE Computational Intelligence Magazine*.

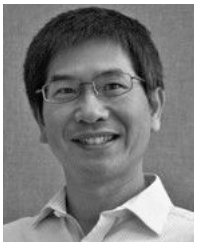
Brent J. Lance (SM'14) received the Ph.D. degree in computer science from the University of Southern California (USC) in 2008. He was with the USC's Institute for Creative Technologies as a Post-Doctoral Researcher, before joining ARL in 2010. He is currently a Research Scientist with the Human Research and Engineering Directorate, U.S. Army Research Laboratory. He was involved in improving the robustness of EEG-based Brain–Computer Interaction through improved integration with autonomous systems. He has authored over 35 technical articles, including a first-author publication on Brain- Computer Interaction in the 100th anniversary edition of the Proceedings of the IEEE.

Dr. Lance is a member of the Association for Computing Machines. He is an Associate Editor of the IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING.

Vernon J. Lawhern received the B.S. degree in applied mathematics from the University of West Florida, Pensacola, FL, USA, in 2005, and the M.S. and Ph.D. degrees in statistics from Florida State University, Tallahassee, FL, USA, in 2008 and 2011, respectively.

He is currently a Mathematical Statistician of the Human Research and Engineering Directorate with the U.S. Army Research Laboratory. He is currently interested in machine learning, statistical signal processing and data mining of large neurophysiological data collections for the development of improved brain–computer interfaces.

Stephen Gordon received the B.S. degrees in mechanical engineering and electrical engineering from Tennessee Technological University in 2003, and the M.S. degree and the Ph.D. degree in electrical engineering from Vanderbilt University. He was involved extensively in the area of brain–computer interfaces. Since 2009, he has been collaborated with researchers at the U.S. Army Research Laboratories Human Research and Engineering Directorate, focused on numerous projects and programs, including the High-Definition Cognition ATO and the Cognition and Neuroergonomics Collaborative Technology Alliance. He is currently a Research Engineer with DCS Corporation, Alexandria, VA. His interests include signal processing and machine learning approaches for neurophysiological analysis and state detection.



Tzyy-Ping Jung (S'91–M'92–SM'06–F'15) received the B.S. degree in electronics engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from Ohio State University, Columbus, OH, USA, in 1989 and 1993, respectively. He is currently a Research Scientist and the Co-Director of the Center for Advanced Neurological Engineering with the Institute of Engineering in Medicine, University of California at San Diego (UCSD), La Jolla, CA, USA, where he is also an Associate Director of the Swartz Center for Computational Neuroscience with the Institute for Neural Computation, and an Adjunct Professor of Bioengineering. In addition, he is an Adjunct Professor of Computer Science with National Chiao Tung University, and an Adjunct Professor of the School of Precision Instrument and Opto-Electronic Engineering, Tianjin University, Tianjin, China. His research interests are in the areas of biomedical signal processing, cognitive neuroscience, machine learning, EEG, functional neuroimaging, and brain–computer interfaces and interactions.

Chin-Teng Lin (F'05) received the B.S. degree from National Chiao-Tung University (NCTU), Taiwan, in 1986, and the master's degree and the Ph.D. degree in electrical engineering from Purdue University, USA, in 1989 and 1992, respectively. He is currently the Distinguished Professor with the Faculty of Engineering and Information Technology, University of Technology Sydney. He also owns Honorary Chair Professorship of Electrical and Computer Engineering, NCTU, International Faculty of University of California at San Diego, and the Honorary Professorship of University of Nottingham. He has authored over 220 journal papers and 97 patents (h-index: 57) in the areas of computational intelligence, fuzzy neural networks, natural cognition, brain-computer interface, intelligent system, multimedia information processing, machine learning, robotics, and intelligent sensing and control, including approximately 108 IEEE journal papers. He is the co-author of *Neural Fuzzy Systems* (Prentice-Hall), and the author of *Neural Fuzzy Control Systems with Structure and Parameter Learning* (World Scientific).

Dr. Lin was elevated to be an IEEE Fellow for his contributions to biologically inspired information systems in 2005, and was elevated International Fuzzy Systems Association Fellow in 2012. He received the IEEE Fuzzy Systems Pioneer Award in 2016, the Outstanding Achievement Award by Asia Pacific Neural Network Assembly in 2013, the Outstanding Electrical and Computer Engineer, Purdue University, in 2011, and the Merit National Science Council Research Fellow Award, Taiwan, in 2009. He served as an Editor-in-Chief of the IEEE TRANSACTIONS ON FUZZY SYSTEMS from 2011 to 2016. He also served on the Board of Governors at the IEEE Circuits and Systems (CAS) Society in 2005–2008, IEEE Systems, Man, Cybernetics (SMC) Society in 2003–2005, the IEEE Computational Intelligence Society (CIS) in 2008–2010, and Chair of the IEEE Taipei Section in 2009–2010. He is the Distinguished Lecturer of the IEEE CAS Society from 2003 to 2005, and CIS Society from 2015 to 2017. He served as the Deputy Editor-in-Chief of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II in 2006–2008. He was the Program Chair of the IEEE International Conference on Systems, Man, and Cybernetics in 2005 and the General Chair of the 2011 IEEE International Conference on Fuzzy Systems.