A Novel Outlier Detection Method for Identifying Torque-related Transient Patterns of *in vivo* Muscle Behavior

Sheng Han, Xin Chen, *EMBS Member*, Sheng Zhong, Yongjin Zhou, *EMBS Member*, and Zhiguo Zhang*, *EMBS Member*

Abstract—**This paper proposed a novel outlier detection method, named l_1-regularized outlier isolation and regression (LOIRE), to examine torque-related transient patterns of *in vivo* muscle behavior from multimodal signals, including electromyography (EMG), mechanomyography (MMG) and ultrasonography (US), during isometric muscle contraction. Eight subjects performed isometric ramp contraction of knee up to 90% of the maximal voluntary contraction, and EMG, MMG and US were simultaneously recorded from the rectus femoris muscle. Five features, including two root mean square amplitudes from EMG and MMG, muscle cross sectional area, muscle thickness and width from US were extracted. Then, local polynomial regression was used to obtain the signal-to-torque relationships and their derivatives. By assuming the signal-to-torque functions are basically quadratic, the LOIRE method is applied to identify transient torque-related patterns of EMG, MMG and US features as outliers of the linear derivative-to-torque functions. The results show that the LOIRE method can effectively reveal transient patterns in the signal-to-torque relationships (for example, sudden changes around 20% MVC can be observed from all features), providing important information about *in vivo* muscle behavior.**

I. INTRODUCTION

In vivo muscle behavior during contraction delivers important information about structure and function of muscles, and is a central topic in neuromuscular research. Some most commonly-used recording techniques developed to measure *in vivo* muscle behavior include electromyography (EMG) [1], mechanomyography (MMG) [2], and ultrasonography (US) [3]. The temporal and spectral characteristics of EMG and MMG, such as its amplitude and spectrum, have been extensively used to examine the motor control strategies [1, 2]. On the other hand, the morphological change of muscle can usually be obtained by US [4, 5]. Since US is noninvasive, real-time and easily accessible, it has been widely applied to measure *in vivo* muscle architecture under both static and dynamic muscle conditions.

Research supported in part by a seed funding (201209160022) from the University of Hong Kong and the National Natural Science Foundation of China (Grant No. 81000637).

Sheng Han and Zhiguo Zhang are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, China email: zgzhang@eee.hku.hk).

Xin Chen, Sheng Zhong, and Yongjin Zhou are with the School of Medicine, Shenzhen University, National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Shenzhen, China.

Multimodal signals (EMG, MMG, or US) are normally recorded as time-series data, but it is often of greater advantage to represent these signals as functions of the torque. It is because the signal-to-torque relationship can provide useful information for understanding how muscle behaviors change with the torque output. Polynomial regression is most widely adopted for the estimation of such signal-to-torque relationship and the results obtained from polynomial regression have shown that EMG, MMG, and US exhibit different response patterns with respect to torque. However, polynomial regression is not flexible enough to describe highly nonlinear and transient relationships between EMG/MMG/US and torque because it is based on a specific functional form (e.g., linear or quadratic) for all observed data. Local polynomial regression (LPR) is recently introduced to present the continuously identified features of three modalities as smooth functions of torque values [6]. Importantly, LPR makes it possible to identify transient response patterns of muscles with the increase of torque from EMG/MMG/US-to-torque smooth functions.

In this study, we proposed to use a novel outlier detection method, named l_1-regularized outlier isolation and regression (LOIRE), to identify outliers in LPR-derived EMG/MMG/US-to-torque relationships, with the aim to examine transient torque-related patterns of EMG, MMG and US signals. We assume that the signal-to-torque functions are basically quadratic functions, and, thus, the first-order derivatives of the signal-to-torque functions (i.e., derivative-to-torque functions) are linear functions. We further assume that the majority of samples in the linear derivative-to-torque functions are contaminated with Gaussian distributed noise, and only a small portion of samples are largely deviated from the ordinary pattern (linear functions) and they can be regarded as transient patterns in the signal-to-torque functions. These transient patterns are represented as outliers in the linear derivative-to-torque functions and can be detected by the proposed outlier detection method. The new LOIRE method is based on l_1 relaxation and is developed for identifying outliers in a linear regression model contaminated with both Gaussian noise and Bernoulli error. The LOIRE method is flexible and efficient, especially when there are a large number of outlier points caused by abrupt changes in the data. Experimental results show that this new outlier detection method can effectively identify torque-related transient patterns in single-to-torque relationships during isometric muscle contraction.

II. Experiments and Methods

A. Signal Acquisition

Multimodal measurements were taken in eight (five males and three females) healthy volunteers (age = 31.4 ± 5.7 years). The study was approved by the local ethics committee. The subject was seated with the right leg at a flexion angle of 90° on a test bench of an isokinetic dynamometer (Humac Norm Testing and Rehabilitation System, Computer Sports Medicine, Inc., MA, USA). The subject was required to put forth his/her maximal effort of isometric knee extension for a period of 6 s. In each test trial the subject was instructed to perform ramp contractions to produce torques increasing linearly from zero up to 90% MVC in 6 seconds. During each contraction, a template torque, serving as a target, and the output of the subject's torque, were displayed simultaneously on a computer screen, which helped the subject to adjust his torque production to track the target torque in real-time. Three trials were repeated with a rest of 5 min between adjacent trials.

The ultrasound images of the rectus femoris (RF) muscle were obtained by an ultrasonic scanner (EUB-8500, Hitachi Medical Corporation, Tokyo, Japan). The ultrasound probe was fixed by a custom-designed multi-degree adjustable bracket. The long axis of the probe was arranged perpendicularly to the long axis of the thigh on its superior aspect, 40% distally from the knee. The ultrasound image was digitized by a video capture card (NI PCI-1411, National Instruments Corporation, Austin, TX, USA) with a frame rate of 25 Hz and resolution of 0.15 mm. Two surface bipolar Ag-AgCl EMG electrodes (Axon System, Inc., NY, USA) were placed on the RF muscle belly parallel with the long axis of the muscle on both sides of ultrasound probe, and a reference EMG electrode was placed near the kneecap. The MMG signal was detected using an accelerometer (EGAS-FS-10-/V05, Measurement Specialties, Inc., France) fixed with two-sided tape. The EMG and MMG signals were amplified by a custom-designed amplifier with a gain of 2000, filtered separately by 10-400 Hz, 5-100 Hz band-pass analog filters, and digitized by a 12-bit data acquisition card (NI-DAQ 6024E, NI, Austin, TX, USA) with a sampling rate of 1 KHz. The isometric torque output from the dynamometer was also sampled by the NI-DAQ card in synchronization with the ultrasound image capture.

B. Feature Extraction

The EMG and MMG signals were segmented into 256-ms epochs. The center of each EMG/MMG epoch was aligned in time with the corresponding ultrasound image according to the timestamp, so that the epochs were registered to the image sequence in time domain. The root mean square (RMS) values of EMG and MMG were calculated for each epoch and expressed as a percentage of their maximal values at 90% MVC. Muscle dimensions were measured offline by in-house image processing program. An object tracking method, named constrained mutual information-based free-form deformation (C-MI-FFD), was used to automatically extract dynamic muscle boundary from the ultrasound image sequence [7]. For each trial, the first image in the sequence was selected as reference and the boundary of the RF muscle was outlined with smooth lines by the investigator. Then the C-MI-FFD method was applied to track the cross sectional area (CSA) boundaries in subsequent images. After getting the boundary, muscle thickness was measured as the greatest vertical distance between the anterior and posterior borders from the extracted boundary, and muscle width was measured at 50% of the vertical distance between the anterior and posterior borders in the boundary, perpendicular to the vertical measure.

C. Local Polynomial Regression

Local polynomial regression (LPR) [8, 9] is a flexible and efficient nonparametric regression method that is particularly effective for interpolation and smoothing of non-uniformly sampled data. Compared with polynomial regression that fits a certain functional form to all observed data, the LPR method is more data-driven and the regression functions are determined locally by windowed data. At any given point of the independent variable, the LPR method fits polynomials to a fraction of data (dependent variable) within a window centered at the point and having variable window bandwidth. The window bandwidth can be adaptively selected by an intersection of confidence intervals (ICI) technique [9] for the best bias-variable tradeoff of the LPR estimator. As a consequence, the LPR method can better capture the dynamic relationship and local information than the conventional polynomial regression. In this study, at each torque sample (uniformly distributed from 0 to 80% MVC with a step of 1% MVC, resulting in 81 torque samples), the LPR used a one-order polynomial (L=1) to fit one set of multimodal features and the local window was adaptively selected by the ICI method. The model order in LPR was set to 1 to avoid possible over-fitting. With the LPR and the ICI methods, the smooth functions of multimodal features and their first-order derivatives with respect to torque were estimated. Subsequently, we use a new outlier detection method to identify outliers from the derivative-to-feature functions for the inference of transient patterns of multimodal signal-to-torque relationships.

D. The LOIRE Method

In this section, we applied the LOIRE method recently proposed in [10, 11] to detect the transient torque-related patterns of muscle behavior. Suppose that the torque-related transient multimodal patterns follow the model:

$$y = ax + c + e + b \tag{1}$$

where y is the samples of the first-order derivatives of feature with respect to torque, x is the torque samples, a (the regression coefficient) and c (a constant) are hidden variables to be estimated, and e and b are Gaussian noise and Bernoulli error, respectively. The non-zero entries in b indicate the outlier points. Then the problem is how to calculate a and c from given samples $\{(x_i, y_i)\}_{i=1,\dots,n}$, where x_i and y_i are the i-th torque level and the i-th first-order derivative of one EMG/MMG/US feature, respectively.

The model (1) assumes that the transient patterns are caused by abrupt changes of muscle behavior and can be modelled as

Bernoulli error. Therefore, the problem can be formulated by the following optimization problem via maximum likelihood estimation:

$$\begin{aligned} &\min_b \|b\|_0 \\ &s.t.\ \ \|\boldsymbol{y} - \boldsymbol{X\alpha} - b\|_2^2 < t \end{aligned} \tag{2}$$

where $\boldsymbol{y} = [y_1, y_2, \cdots, y_n]^T$, $\boldsymbol{X} = \begin{pmatrix} x_1 & \cdots & x_n \\ 1 & \cdots & 1 \end{pmatrix}^T$, $\boldsymbol{\alpha} = [a, c]^T$, and t is a constraint parameter. In (2), we actually bounded the norm of estimated Gaussian noise by the parameter t, which is acceptable since large-magnitude noise occurs with a very small probability and can be regarded as outliers that will be presented in b.

However, Eq. (2) is an NP-hard problem that is almost unsolvable. In most cases, (2) can be relaxed as the following formulation

$$\begin{aligned} &\min_b \|b\|_1 \\ &s.t.\ \ \|\boldsymbol{y} - \boldsymbol{X\alpha} - b\|_2^2 < t. \end{aligned} \tag{3}$$

Eq. (3) is a tractable problem and can be solved efficiently via convex optimization. Without loss of generality and for simplicity, we transform (3) to the following formulation via the Lagrange Dual theory [12]:

$$\min_b \|b\|_1 + \lambda \|\boldsymbol{y} - \boldsymbol{X\alpha} - b\|_2^2 \tag{4}$$

where λ is a penalized coefficient determined by t. In (4), without loss of generality, we suppose $\boldsymbol{X}$ is a full rank matrix (if not, singular value decomposition can be used to decompose matrix $\boldsymbol{X}$ to derive a full rank matrix). An efficient algorithm for (4) is given in Table 1.

TABLE 1
OUTLIER DETECTION ALGORITHM FOR (4)

Given $\boldsymbol{y}$ and $\boldsymbol{X}$

Initialization: $k = 0, b_0 = \vec{0}$

WHILE $\| b_k - b_{k+1} \|_1 > \varepsilon$ **DO**

$\alpha_{k+1} = (X^T X)^{-1} X^T (y - b_k)$

$y_{k+1} = y - X\alpha_{k+1}$.

$b_{k+1} = sign[y_{k+1}] \circ \left[|y_{k+1}| - \frac{1}{\lambda} \vec{1} \right]_+$

$k = k + 1$

END WHILE

OUTPUT b

The operator "$\circ$" denotes a Hadamard product (i.e., if $z = x \circ y$, then z_i, the i-th coordinate of vector z, equals to $x_i y_i$). For the notation "$[\cdot]_+$", suppose $p = [q]_+$, we have $p_i = q_i$ if $q_i > 0$, otherwise $p_i = 0$. The vector $\vec{1}$ is a vector with a value of one at each entry.

Due to page limitation, the convergence proof of the algorithm is omitted here. Interested readers can refer to [10,11] for analysis details. After the above outlier detection step, entries of y_i with the corresponding are identified as outlier. In order to achieve higher accuracies for parameters a and c in (1), we can, with some theoretical supports from paper [10,11], use the classical Least Mean Squares/Minimum Mean Squares [13] regression on the cleaned data y with its detected outlier entries being removed.

In this paper, we applied the outlier detection model in (4) and the algorithm in Table 1 to identify outliers in the EMG/MMG/US-to-torque derivative functions, which are averaged across eight subjects. The detected outlier samples are presented as non-zero entries in the Bernoulli error b and are regarded as torque-related transient response patterns of this group of people.

III. RESULTS

The EMG, MMG and torque signals during one representative trial of one subject are shown in Figure 1. It can be clearly seen that both RMS_{EMG} and RMS_{MMG} are increasing functions of time, while torque almost linearly increases with respect to time.

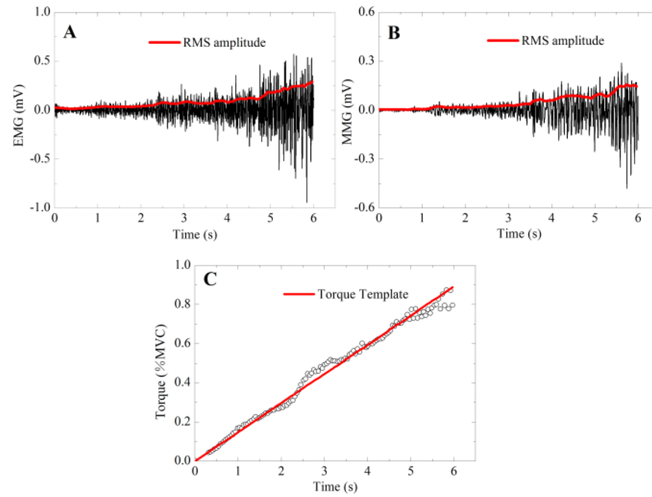

Figure 1. The time courses of (a) EMG, (b) MMG, and (c) torque during a representative trial. The torque signal was overlaid onto the ramp template as it appeared for the subject during the trial. The RMS_{EMG} and RMS_{MMG} curves are also overlapped in the raw EMG/MMG curve.

The dimensional changes of CSA, thickness and width of this trial are shown in Figure 2. We can see that all three features are highly nonlinear with respect to time and they provide complementary morphological information regarding the muscle architecture.

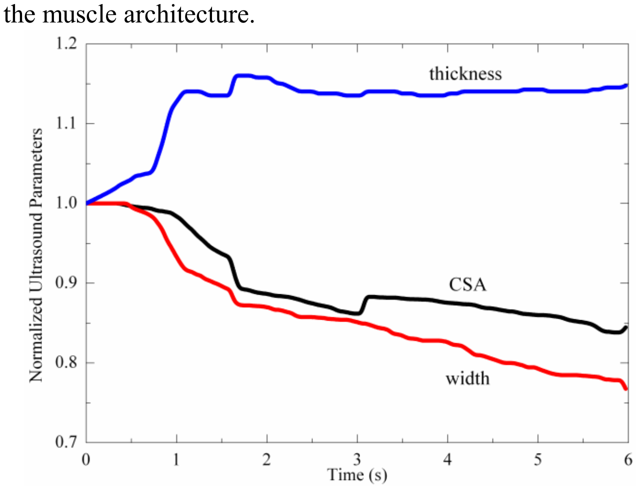

Figure 2. Dimensional changes of CSA, thickness and width in a representative US trial. These US features are automatically extracted by the C-MI-FFD method [7].

Figure 3 shows the LPR-derived smooth functions and derivative functions of five features as functions of the torque, which are all averaged from eight subjects. We can see that, at the group level, all these signal-to-torque relationships exhibit dynamic and transient patterns. The linear regression functions (which describe the overall relationship between feature derivatives and torque) and the transient patterns estimated by the LOIRE method are respectively indicated by dashed black lines and gray background in the bottom panel of Figure 3. Taking CSA as an example, we can see that the CSA-to-torque function has a negative quadratic coefficient and around 20% MVC there is a sudden decrease in the derivative-to-torque function of CSA, which means CSA rapidly decreases with torque around 20% MVC in a significantly different decreasing rate from that in other ranges of torque. Specifically, a sudden change around 20% MVC can be observed from all features, and is particularly pronounced for US features.

IV. Discussion and Conclusion

The relationship between muscle activities and generated torque during muscle contraction is an important and challenging topic. The most common approach to characterize the relationship is polynomial regression, but polynomial regression is not flexible enough to describe transient patterns in the whole torque spectrum. This paper introduced a new outlier detection method, LOIRE, to detect transient patterns of *in vivo* muscle behavior during isometric contraction. Based on the assumption that the overall relationship between muscles signals and torque is quadratic (so the overall derivative-to-torque relationship is linear), the proposed outlier detection method is used on derivative-to-torque functions, which are estimated using local polynomial regression, to detect outliers as transient patterns of muscle. For the first time, dominant transient patterns around 20% MVC in all multimodal muscle features are quantitatively identified, though this critical point of 20% MVC has been reported in literature. The LOIRE method (as well as the whole data analysis pipeline) can provide novel and more complete information of muscle contraction, and is potentially a useful tool for the muscle assessment in clinical applications.

References

[1] J. V. Basmajian, and C. J. DeLuca, *Muscles Alive: Their Functions Revealed by Electromyography*, 5th ed., Baltimore : Williams & Wilkins, 1985.

[2] C. Orizio, "Muscle sound: Bases for the introduction of a mechanomyographic signal in muscle studies," *Crit. Rev. Biomed. Eng.,* vol. 21, no. 3, pp. 201-243, 1993.

[3] M. V. Narici, T. Binzoni, E. Hiltbrand, J. Fasel, F. Terrier, and P. Cerretelli, "*In vivo* human gastrocnemius architecture with changing joint angle at rest and during graded isometric contraction," *J. Physiol.-London,* vol. 496, no. 1, pp. 287-297, Oct, 1996.

[4] P. Augat, and F. Eckstein, "Quantitative imaging of musculoskeletal tissue," *Annu. Rev. Biomed. Eng.,* vol. 10, pp. 369-390, 2008.

[5] S. Bianchi, and C. Martinoli, *Ultrasound of the Musculoskeletal System*: Berlin, New York :Springer, 2007.

[6] X. Chen, S. Zhong, Y. Y. Niu, S. P. Chen, T. F. Wang, S. C. Chan, and Z. G. Zhang, "A multimodal investigation of in vivo muscle behavior: System design and data analysis," *ISCAS2014*, Melbourne, Australia, 1-5 Jun, 2014.

[7] X. Chen, Y. P. Zheng, J. Y. Guo, Z. Y. Zhu, S. C. Chan, and Z. G. Zhang, "Sonomyographic responses during voluntary isometric ramp contraction of the human rectus femoris muscle," *Eur. J. Appl. Physiol.,* vol. 112, no. 7, pp. 2603-2614, Jul, 2012.

[8] V. Katkovnik, K. Egiazarian, and J. Astola, *Local Approximation Techniques in Signal and Image Processing*, Bellingham, WA: SPIE Press, 2006.

[9] Z. G. Zhang, S. C. Chan, K. L. Ho, and K. C. Ho, "On bandwidth selection in local polynomial regression analysis and its application to multi-resolution analysis of non-uniform data," *J. Signal Process. Sys,* vol. 52, no. 3, pp. 263-280, 2008.

[10] S. Wang, S. Han, and W. S. Wong, "Efficient robust regression via l_1 relaxation", *submitted to NIPS2014*.

[11] S. Han, S. Wang, D. Tao, and X. Wu, "*l_1-Regularized Outlier Isolation and Regression*", *arXiv: 1406.0156*.

[12] S. Boyed and L. Vandenberghe, "Convex Optimization," Cambridge University Press, Cambridge, 2004, pp. 215-231.

[13] L. L. Scharf and C. Demeure, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Reading, MA: Addison-Wesley, 1991.

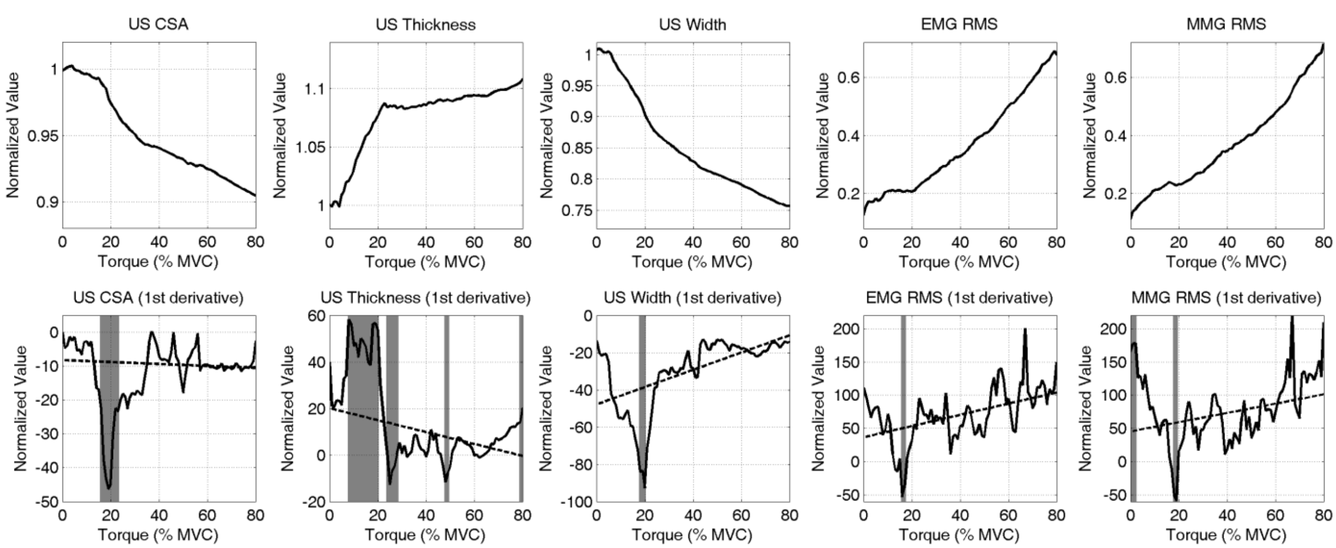

Figure 3. LPR on five parameters as functions of torque. Top: smooth functions; bottom: first-order derivative functions. The bottom panel also indicates the transient patterns detected by the proposed LOIRE method with gray background.The bashed black lines are the linear regression functions (determiend by *a* and *c*) estimated by the LOIRE method.