Exploring Possibilities for Real-Time Muscle Dynamics State Estimation from EMG signals

Luciano L. Menegaldo

Abstract—Real-time estimation of muscle state (activation, force and length) can be used for visualization of muscle function in neuromuscular rehabilitation and control of wearable assistive and augmenting devices. This paper proposes two formulations for developing linear state estimations of muscle dynamics in isometric contractions: the Luemberger observer and the Kalman filter. One of the main requirements of an estimator for such application is small lag, and a Finite Impulse Response (FIR) pre-filter was used to eliminate part of the high-frequency content of the EMG signal. A run test with human soleus EMG sample was performed. Both Luemberger observer and Kalman filter provided reasonably accurate estimations of the muscle state, compared to usual off-line EMG-driven analysis. In addition, the role of some Kalman filter parameters is analyzed.

I. INTRODUCTION

Muscle dynamics is an active research topic in several fields, including: biomechanics, sports science, human and animal physiology, orthopedics, rehabilitation etc. In the emerging field of biomechatronics, the ability to model, simulate and control this dynamics is paramount for many applications. Controlling a haptic device or some wearable assistive system can be performed by using several approaches, from classical PID controllers to artificial intelligence neuro-fuzzy systems.

Modeling muscle dynamic behavior in the complete dynamic system is important to formulate, simulate and implement the control strategy. A wide spectrum of muscle dynamics formulations can be found in biomechanics literature, from simple algebraic equations to very complex molecular-level dynamic models. The most explored class of models is based on the seminal work of Hill [1], that gave origin to the so-called “Hill-type muscle models”. These models consist of an association of contractile, elastic and damping elements, leading to systems of a few non-linear ordinary differential equations (ODEs).

Our group has been using a particular formulation of a Hill-type muscle model [2], based on the work of Zajac [3], for estimating muscle forces from EMGs [2], [4], [5]. This approach is known as EMG-Driven force estimation [6], [7]. The basic idea is to estimate muscle forces from EMG data, considering the rectified and low-pass filtered EMG as the excitation input for the dynamic muscle model, that is then integrated numerically off-line. For real-time (RT) applications, the key problem becomes the EMG signal processing, specially filter lags, and model numerical integration processing time. EMG is a complex signal and EMG-driven models require a very smooth low-frequency envelope as input signal. A digital filter frequently used for this task is the 6th order low-pass Butterworth, with design cut-off frequency of 2 Hz. This element introduces a large delay (see Sec. III), that is easily compensated by forward and reversed filtering [8]. For RT applications, this approach is no longer suitable.

Depending on the application, a specific maximum allowed lag is expected to be required. A possible use for RT EMG-driven models is aiding neuromuscular rehabilitation following a orthopedic surgery, such as an Achilles tendon reconstruction. In this case, the healing process is highly sensitive to the loading levels, and EMG-driven models can give to the patient and to the therapist a RT biofeedback of the muscle force [9]. According to [9], the maximum allowed lag between two frames for still providing a RT visual sensation of a computer-generated virtual limb is around 40 ms.

In the case of controlling biomechatronic devices, the lag is directly associated with control stability margins (phase margins), as known from classical linear control theory. Depending on system dynamics and control synthesis method, it can be more or less robust to sensor and processing lags.

In this paper we explore some possibilities for using alternative filters with lower-delay associated with state observers, to provide fast and reliable estimations of the complete muscle state. In our case, the state vector comprises 3 state variables: muscle activation, muscle force and muscle length (of the contractile part). This model has two control inputs: musculotendon length and neuromuscular excitation. Since only isometric contractions will be addressed here, excitation input will be considered as the control, while musculotendon length remains a fixed parameter. By hypothesis, the only available sensor signal is the EMG. Two state estimators will be tested: the classical Luemberger Observer (LO/LQR), with gains found by a Linear Quadratic Regulator (LQR) and the Kalman filter (KF).

II. MUSCLE MODEL LINEARIZATION

The details of the musculotendon actuator model has been described in [2] and [5]. Essentially, it comprises a set of three elements: contractile (C), elastic ($k^{PE}$) and damping elements (B) in series with a elastic linear tendon ($k^{T}$) (Fig. 1).

The model is described by three ODEs (eq. 1): activation dynamics, contraction dynamics and an auxiliary equation stating that the muscle velocity is the derivative of muscle
length. This variable is important due to the dependency of the muscle force generating capacity with muscle length [3], the force-length relationship. The function \( g \) is an lengthy algebraic expression that incorporates force \( \times \) muscle velocity and force \( \times \) muscle length relationships (see details in [2]).

\[
f(x, u) = \begin{bmatrix} \dot{a} \\ F^T \\ L^M \end{bmatrix} = \begin{bmatrix} (u-a)(k_1u + k_2) \\ g(F^T, a, u, L^M) \\ \dot{V} = dL^M/dt \end{bmatrix}
\]

(1)

In these expressions, \( u \) is the neuromuscular excitation (input variable), \( a \) is activation, \( k_1 \) and \( k_2 \) are constants, \( F^T \) is the tendon force, \( L^M \) is the muscle length and \( V^M \) the muscle velocity. The model has been formulated dimensionless, thus the state and control variables are all contained into the interval \([0, 1]\): 0 no activity, 1 fully activated. Force is normalized by the maximum muscle force, and length by the optimal length, i.e., the muscle length which the muscle is able to exert maximum force [3]. The control variable \( u \in [0, 1] \) is an overall measure of the muscle input, comprising both motor units summation and action potential frequency modulation effects. Simulation parameters were chosen for soleus human muscle, using data from Delp’s OpenSim lower limb model [10]. The state variables vector is defined as:

\[
x = \begin{bmatrix} a \\ F^T \\ L^M \end{bmatrix}
\]

(2)

Initially, we have tried to linearize this model by the Jacobian method, calculating the Jacobian matrices \( \frac{df}{dx} \) and \( \frac{df}{du} \) calculated at several operating points. The response resulting from the linearized state-space integration did not gave accurate results compared to the non-linear model, in special for muscle force and length variables.

A second attempt was identifying a linear model by minimal-squares method [11] from the results obtained by the integrated non-linear system. The procedure can be summarized as follows:

1) Generate a frequency-rich input signal \( u = [u_i] \), \( i = 1..N \), \( N \) is the number of control points, such as a step added with a slightly low-pass filtered Gaussian noise;

2) Integrate numerically the ODEs with \( u \) in (1) to find a sequence \( [X] = [x(1), x(2), x(3), \ldots]_{\text{Ntimes}} \) containing the state trajectories, where \( x(1) \) is the activation etc. Generate a \( [\alpha] = [x(1), x(2), x(3), u]_{\text{Ntimes}} \) matrix;

3) Calculate the derivatives of the \( [X] \) trajectory, \( \dot{X} \), using the function where the non-linear model was defined and that was used in integration of step 1. Write these derivatives in a matrix \( [\beta] = X(1), X(2), X(3), \ldots]_{\text{Ntimes}} \);

4) Solve the expression \( [\gamma]_{\text{A3}} = ([\alpha] [\alpha]^+ [\alpha] [\beta] \), where \( + \) is the Moore-Penrose pseudo-inverse matrix.

5) The terms \( \gamma(1, 1), \gamma(1, 2), \gamma(1, 3), \gamma(2, 1), \ldots, \gamma(3, 3) \) forms the \([A] \) linear state matrix and the terms \( \gamma(4, 1), \gamma(4, 2), \gamma(4, 3) \) the control matrix.

Figure 2 shows a sample result comparing the integrated non-linear and linearized equations for a noise-added step function.

III. EMG ACQUISITION AND FILTERING

An EMG sample of human soleus muscle performing a maximum voluntary contraction was used in the analysis. The signal was sampled at 2 kHz, and all details concerning the subjects, ethics committee, acquisition methodology has been described in [2] and [5]. The task consisted of a 5 seconds sustained Maximum Voluntary Contraction (MVC), preceded and followed by relaxing periods. The raw EMG sample was rectified and normalized by the maximum filtered EMG amplitude for the same sample.

The low-pass filtered that is normally used for extracting the envelope \( u(t) \) in off-line EMG-driven analysis is the 6th order Butterworth IIR filter, 2Hz cut-off frequency [2], that can be considered as our current “gold standard”. However, if applied only in the forward direction, introduces a large phase distortion into the filtered signal, which group delay is shown in Fig. 3. Since the acquisition is being performed at 2kHz, the delay may reach 500ms in some frequencies, what is unacceptable for RT applications.
A more appropriate filter for this purpose is the Finite Impulse Response (FIR). We have designed it using matlab function \texttt{firl.m} \cite{12}. It can be observed, in Fig. 3 that the lag is constant along the frequency spectrum, as expected for this kind of filter, corresponding to half of the filter order. For this 150th order filter, the group delay of 75 points corresponds to a 37.5ms lag. Figure 4 shows the sample EMG signal processed by different ways: rectified and amplitude normalized only (raw), IIR filtered, IIR filtered forward and backward and FIR filtered only forward. It is possible to observe that the FIR filter can significantly attenuate most of the high-frequency content of the signal without visually observable phase distortion.

IV. FULL-ORDER LUENBERGER STATE OBSERVER

State observers are normally used in state space based control systems, where only some of the states are measured by sensors. The remaining states are estimated using the system model. Let \( A, B, C \) and \( D \) be the matrices that describe a linear state space system with state vector \( x \) and output vector \( y \). A new set of observed or estimated state variables \( \hat{x} \) is generated. The observer output is \( \hat{y} = [C] \hat{x} \), where \([C]\) is the output matrix that is subtracted from the real system output \( y \). This difference is summed into the observer dynamical equations and minimized through feedback, forming eq.3, the Luemberger observer.

\[
\dot{\hat{x}} = A \hat{x} + Bu + L (y - C \hat{x}) \quad (3)
\]

\[
\dot{\hat{x}} = (A - LC) \hat{x} + Bu + Ly \quad (4)
\]

\( L \) is a gain vector, that can be designed by some pole placement or LQR technique, such that the error \( \hat{x} = \hat{x} - x \) is minimized.

For formulating a state observer, the linear system must be observable. This property depends on observation matrix \([C]\), used to calculate the observability matrix. The key point here is to decide what information can be used to generate \([C]\) and thus the measurement \( y \). For a non isometric task, maybe joint kinematics could be used for this purpose. Here, we are addressing an isometric task and this information cannot be used. The only available information is the input \( u \) derived from the EMG. Since activation dynamics is described by a simple ODE (eq. 1) and depends on \( u \), we integrate eq. (5) simultaneously with the observer, using Euler method:

\[
a_k = a_{k-1} + \Delta (u_k - a_{k-1}) (k_1 u_k + k_2) \quad (5)
\]

The LQR problem was solved for the observer gain \( L \) defining the following \( Q \) and \( R \) matrices, by trial and error:

\[
Q = \text{diag}([1 \quad 0.001 \quad 0.0001]) \quad (6)
\]

\[
R = 1 \quad (7)
\]

Observation matrix is defined as \( C = [1 \ 0 \ 0] \). Equation (4) can be integrated and compared to the non-linear Runge-Kutta integrated non-linear dynamics. (Fig. 5). The response of the non-linear system (eq. 1) integrated with the forward and reverse IIR filter, will be considered the nominal response, corresponding to the best possible estimate. There is almost no differences between the system and observer responses. In addition, no systematic errors seem to be present in the FIR filtered responses, oscillating around the nominal curve. Very similar results can be obtained when the linearized (sec. II) model is integrated together with the observer, spending however a shorter computational time.
V. KALMAN FILTER ESTIMATION

Kalman filter is a recursive observer formulation that allows finding optimal state estimates, in the minimum variance least squares sense, for a discrete linear system, assuming that the dynamic process and measurements are contaminated with Gaussian noise. Process noise must be uncorrelated with the measurements and both independent of the input.

In our case, these conditions are not fulfilled, since the muscle model is nonlinear and the EMG presents a complex band-limited frequency spectrum, that cannot be considered as a white noise [13]. Some authors observed that whitening of EMG signals can improve amplitude estimation [14], when simple estimators such as RMS and ARV (Average Rectified Value) are used. The whitening problem can be interpreted as filtering the signal with the inverse of the AR (autoregressive) linear stochastic process, obtained through spectral estimation, that fits an EMG sample, providing approximately a signal with white noise spectral characteristics [13]. The whitening filter is therefore MA (moving average) all-zeros filter. We have made some preliminary tests applying a simple whitening filter, designed with Yule-Walker method (matlab function aryule.m), to the raw EMG data before running the KF. No remarkable improvement in the estimations were observed in the results. Since the EMG must be rectified to make sense to the dynamic model, the signal has no longer Gaussian distribution characteristics. In any case, this problem needs further studies and, for now, no whitening was applied at all.

In Kalman Filter equations (eqs. (8)-(12)), $A_{[3x3]}$ and $B_{[3x1]}$ are the muscle state-space matrices discretized using zero-order hold approximation. The output matrix $C = [1 \ 0 \ 0]$ assumes that the available external measurement is activation. The vector sample $y_k$ is the external measurement, $\nu$ is the sensor noise, $k$ is the sample number, $x_k$ is the state vector, $u_k$ the input vector and $w$ is the process noise. An estimate $\hat{x}_k$ of the state vector is found by a Kalman Filter:

\begin{equation}
\hat{x}_k^- = A\hat{x}_{k-1}^- + B u_k \tag{8}
\end{equation}

\begin{equation}
P_k^- = A P_{k-1}^- A^T + Q \tag{9}
\end{equation}

\begin{equation}
K_k = P_k^- H^T (H P_k^- H^T + R)^{-1} \tag{10}
\end{equation}

\begin{equation}
\hat{x}_k = \hat{x}_k^- + K_k (y_k - C \hat{x}_k^-) \tag{11}
\end{equation}

\begin{equation}
P_k = (I - K_k C) P_k^- \tag{12}
\end{equation}

where $P$ is the state estimate error covariance matrix, $R$ is the sensor noise covariance matrix, $Q$ is the process noise matrix and $K_k$ is the Kalman gain.

Vector $y_k$ is expected to contain Gaussian noise $\nu$ and do not need to be sampled at the same frequency of the input. The first choice was estimating the external activation measurements through eq. (5), as shown in Sec. IV. Again, an optimality condition for the KF is not met: the externally measured activation is directly driven by the input, i.e., the processed EMG signal, and therefore autocorrelated with it. In such cases, a possible solution is using the derivative of $y$ as the external measurement [15].

We have tested a strategy for reducing external measurement noise by updating the Kalman gain at a smaller frequency compared to the prediction step (decimation). This kind of situation is frequently found in inertial localization applications of KFs, when the external measurements are provided by a GPS (Global Positioning System), working at a much smaller frequency compared to other inertial sensors. In the interval between two updates, $y_k$ samples are accumulated and, in the end, have their average value evaluated, working as an average filter. This procedure introduces a delay effect in the external measurement. It was observed a decrease in the correlation between $y$ and raw EMG input $u$, from 0.7081 to 0.6286. In the meanwhile, the KF predicts the estimated states, delivering an observed state-space vector at full EMG acquisition frequency, what is convenient for control purposes.

For applying the KF to predict the muscle states, some parameters must be adjusted, specially the $Q$ and $R$ covariance matrices. As discussed before, there are also several options for pre-filtering the raw EMG signals, eliminating part of the noise. These filters are subject to a trade-off between noise attenuation effectiveness and phase lag. For the covariance matrices, it was observed, by trial and error, that the following values can provide correct amplitude estimations for all states:

\begin{equation}
Q = diag([100 \ 0.1 \ 0.1]) \tag{13}
\end{equation}

\begin{equation}
R = 1 \tag{14}
\end{equation}

As an attempt to explore the above mentioned possibilities, five simulations were performed, which conditions
are summarized in Table I. Figures 6, 8 and 10 show the KF estimated activation, tendon force and muscle length, respectively. In Figures 7 and 9, it is possible to observe zoomed views of activation and tendon force, such that the reader may have a better idea of the waveforms. The result of a usual off-line EMG-driven analysis, integrating the non-linear EDOs, is also shown, for evaluating the estimation accuracy.

VI. DISCUSSION AND CONCLUSIONS

This paper is an introductory exploration regarding the formulation of state-space observers for muscle dynamics, using only EMG signals as inputs. Its two main application fields are real-time visualization of muscle state and control of assistive wearable devices. A non-linear model of muscle dynamics was linearized and two linear state observer formulations were tested: Luenberger observer and Kalman filter. Such non-linear models of muscle dynamics are usually used in off-line EMG-driven muscle force estimation analysis, integrating the ODEs with zero-lag IIR forward and backward filtering of the EMG inputs. For RT, this approach is no longer valid, due to computational and filtering delays. One of the possibilities explored was pre-filtering the raw EMG signal with a high-order FIR filter.

The linearization procedure, based on a leasts squares identification of a linear model of the the system subjected to a frequency and amplitude rich input signal provided an accurate linear representation of the non-linear muscle model (Fig. 2).

The Luenberger observer has shown to be adequate for providing a correct estimation of the state variables (Fig. 5). In a feedback control application, integrated to a broader control system scheme, it could possibly improve robustness, compared to estimation of the state variables through direct integration of the ODEs [16].

Regarding KF performance, it has reduced part of the
undesired high-frequency content and provided accurate estimations of the states. Using a pre-filter (e.g. FIR) has improved smoothness of the estimations, although some fast but attenuated fluctuations are still present. Decimation and averaging of the external measurements in the update phase has show to efficiently reduce noise, at the cost of increasing the delay.

The activation variable was less affected by high-frequency KF attenuation, and decimation has shown to be more efficient than FIR pre-filtering (Figs. 6 and 7). Comparing the wave patterns resulting from both, decimation plus averaging reveals a smaller amplitude and higher frequency content than FIR pre-filtering (Figs. 7 and 9). In tendon force and muscle length, significantly smoother estimations were obtained. Depending on the particular application requirements, regarding stability robustness to delays and noise, one of the methods presented can be selected: case 4 presents smaller delays and case 5 more noise attenuation.

Further work is needed to systematically explore the role of the many parameters necessary to formulate the problem: pre-filtering formulation and alternate versions of the KF, to deal with colored noise and model non-linearities (e.g. Extended Kalman Filter). Tests on additional EMG samples and other muscles are necessary. Non-isometric conditions will provide interesting alternative ways for providing uncorrelated y vectors, such as joint kinematics.

VII. ACKNOWLEDGMENTS

The authors are gratefully acknowledged to CAPES (Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior), CNPq (Conselho Nacional de Desenvolvimento Cientifico e Tecnologico) and FAPERJ (Fundacao de Amparo a Pesquisa do Estado do Rio de Janeiro) for financial support.

REFERENCES