Smoothness as a failure mode of Bayesian mixture models in brain-machine interfaces

Siamak Yousefi, Alex Wein, Kevin Kowalski, Andrew Richardson, Lakshminarayan Srinivasan

Abstract— Various recursive Bayesian filters based on reach state equations (RSE) have been proposed to convert neural signals into reaching movements in brain-machine interfaces. When the target is known, RSE produce exquisitely smooth trajectories relative to the random walk prior in the basic Kalman filter. More realistically, the target is unknown, and gaze analysis or other side information is expected to provide a discrete set of potential targets. In anticipation of this scenario, various groups have implemented RSE-based mixture (hybrid) models, which define a discrete random variable to represent target identity. While principled, this approach sacrifices the smoothness of RSE with known targets. This paper combines empirical spiking data from primary motor cortex and mathematical analysis to explain this loss in performance. We focus on angular velocity as a meaningful and convenient measure of smoothness. Our results demonstrate that angular velocity in the trajectory is approximately proportional to change in target probability. The constant of proportionality equals the difference in heading between parallel filters from the two most probable targets, suggesting a smoothness benefit to more narrowly spaced targets. Simulation confirms that measures to smooth the data likelihood also improve the smoothness of hybrid trajectories, including increased ensemble size and uniformity in preferred directions. We speculate that closed-loop training or neuronal subset selection could be used to shape the user’s tuning curves towards this end.

Index Terms—Bayesian mixture model, brain-machine interface, general purpose filter design, neural decoding, neuroprosthesis.

I. INTRODUCTION

A core challenge of the brain-machine interface (BMI) is to map (or decode) neural signals into smooth trajectories of an assistive device that reflects the user's intentions. This paper focuses specifically on Bayesian mixture models used to drive reaching movements to one of multiple potential discrete targets [2, 3] that can be identified from computer vision [4] or neural recordings [5, 6].

These mixture models incorporate reach state equations (RSE), which are stochastic difference equations that specify the statistical dependence between target and path in a trajectory [2, 3, 7-14]. RSE-based mixture models (MM-RSE) improve the estimation of reaching movements from neural signals, as demonstrated in simulation [2] and with non-human primate data [3, 9]. In this paper, we explore the relationship between MM-RSE and the smoothness of reach trajectories.
Figure 2. Basic tuning curve properties. Tuning curves for (A) each of 12 neurons in subject 1 and (B) each of 11 neurons in subject 2, with preferred directions and target locations indicated as points on the x-axis. (C, D) Alternate representation of preferred directions and target locations, arranged according to the 2D workspace for (C) subject 1 and (D) subject 2. Targets were positioned on a circle with 10 cm radius.

A. Related Work

1. Bayesian BMI mappings

MM-RSE fall within the generic class of recursive Bayesian filters [15]. In the BMI setting, the latent variable model (or state equation) of the filter represents the user’s evolving intention for the device. The observation model of the filter represents the probability distribution of neural activity conditioned on user intent. The filter incorporates the latent variable model and observation model into equations for calculating a continuously updated estimate of user intent from the latest recorded neural activity.

Although BMI mappings like standard linear regression and population vector analysis [16] have been more widely implemented in animal [17-19] and human [20, 21], Bayesian estimation can generate smoother trajectories of the assistive device [22, 23]. This occurs because Bayesian estimation employs a state equation, which promotes smooth output, and represents a prior (in the statistics sense) on the expected evolution of the user’s intention.

Developments in Bayesian mappings for BMI applications are divided into three categories. First, various procedures for calculating and representing the posterior probability density have been explored. These include the particle filter [24], approximate point process filters [25], and others. Second, various neural signal models have been described. Variants include Gaussian observation models [26], point process models [27], and history-dependent observation models [27], where this list of variants is not comprehensive. Third, various state equations (latent variable models) have been examined, corresponding to different priors on the expected range of user intentions [2, 3, 7, 9-13]. The subject of our investigation is this third class of developments.

2. Reach state equations and mixture models

Latent variable models called reach state equations (RSE) were previously developed to improve the quality of reaching movements generated by brain machine interfaces [2, 3, 7, 9-14, 28]. Data-driven RSE require a database of example reaching movements to a set of target locations [3, 9]. Generative RSE provide a formulation for the state equation that incorporates arbitrary target locations without the need for a training database [2, 7, 10-14]. When the target identity is known with certainty, or near certainty, the resulting Bayesian mapping [7] generates smooth trajectories, a substantial improvement over the random walk latent variable model used in the standard Kalman filter [29].

The RSE was subsequently extended using mixture models to support multiple potential reach targets [2, 3]. Mixture models have also been applied in BMI research for recursive spike sorting [30, 31]. In the reach application, the generic MM-RSE processes neural signals in four steps (Figure 1A). First, spikes are binned into a vector containing spike counts for all neurons in the ensemble. Second, RSE-based filter output is computed in parallel for each possible target. Each of these parallel filters uses a RSE that corresponds to its specific target. Third, the probabilities of intent for each target are updated. Fourth, the parallel filter outputs are averaged, weighted by the target probabilities. While details of implementation differ between the various MM-RSE, these four steps are the conceptual basis for our subsequent analysis of smoothness. The mathematical analysis in this paper (sections 3.4-3.5) operates at this general level of description in examining the smoothness of MM-RSE. The empirical (sections 3.1-3.3) and simulated (section 3.6) analysis in this paper uses the specific implementation of MM-RSE described in [2] and the RSE described in [7, 32] to confirm the more general mathematical analysis.

3. Smoothness in open-loop versus closed-loop evaluation

Related work has examined trajectory smoothness as a measure of performance in closed- and open-loop validation [23, 33-35]. In closed-loop validation, including the target clinical application, the subject observes and adjusts to the output of the BMI mapping. In open-loop validation, neural activity is recorded in the presence of overt arm movements, and feedback about BMI output is not available to the user. This paper focuses on smoothness in open-loop validation of MM-RSE.

Prior work on trajectory smoothness [23, 33-35] provides justification for why analysis of smoothness in the open-loop setting might translate into closed-loop validation. In the context of cursor control, algorithmic differences related to the velocity trajectory smoothing can impact both open- and closed-loop performance [23], even when closed loop performance diminishes certain decoding biases [23, 33]. Separate work examined mean integrated distance to target
(MID) as a measure of trajectory position smoothness, showing that smoothness can substantially improve with closed loop control [34]. Subsequent theoretical work using MID in a stochastic control model for the brain in closed-loop BMI operation [35] showed that although closed-loop performance improves MID overall, worse decoding performance in open-loop still translates into worse MID-related performance in closed-loop, providing alternate theoretical confirmation of [23] and first order trends in [34].

B. Contributions of This Manuscript

Understanding if and why mixture models degrade RSE performance is a basic step towards engineering BMI that support reliable, high-performance reaching movements. Despite its strength in decoding accuracy, we hypothesize that MM-RSE result in significantly less smooth trajectories than RSE-based mapping to known targets. This hypothesis was qualitatively suggested in a prior publication [2]. The goal of this paper is to quantitatively document and explain this degradation in performance using empirical data and mathematical analysis of MM-RSE. We also provide preliminary recommendations for improvement based on simulation. Because the mixture model is a Bayesian approach, this analysis is expected to be particularly relevant when neural signals alone leave substantial uncertainty in user intent. Example settings include BMI, both invasive and non-invasive, with limited neural signal channel counts or large signal variability relative to user intent.

II. METHODS

A. Measure of Smoothness

We use average magnitude of angular velocity (degrees/ms) as a meaningful and well-defined measure of smoothness (Figure 1B). This quantity represents the minimum angle subtending two consecutive decoded velocity vectors. Under this measure, a perfectly goal-directed movement would have zero angular velocity at each time step, although the speed of movement would rise and fall. This notion of smoothness is related to the sliding window average of velocity estimates used to smooth output from the population vector algorithm (PVA) and other methods [33]. The sliding window average serves to decrease the Euclidean difference between consecutive velocity estimates, typically decreasing changes in magnitude of angular velocity in movement direction (our measure of smoothness).

Natural arm trajectories may not always be smooth by this measure, such as when the arm overshoots a target, generating significant angular velocity during corrective behavior. If the BMI allows the user to avoid needing to make corrective behaviors, it is possible that BMI performance could exceed natural arm movements by this measure. The possibility of exceeding natural performance with an assistive device does not invalidate this measure. By such logic, speed would be invalidated as a measure of performance in a 50 meter sprint for amputee runners with enhanced prosthetic legs. A BMI that allowed for smoother movements than possible with natural limb movement would be a desirable advance.

B. Simulation Procedure

Portions of this analysis were conducted with open-loop simulation as follows. Intended center-out reaching movements in a 2-dimensional workspace were simulated to one of eight randomly chosen targets arranged on a circle using the RSE [7, 32]. Ensemble spiking activity was simulated using the time rescaling theorem [36] based on the intended movement trajectory using a conventional cosine-tuned point process model developed previously [27, 36, 37], provided in equation (1) below, and applied extensively in our prior work on RSE and MM-RSE [2, 7, 13].

Under the basic discrete-time model, the instantaneous probability of spiking at time step \( k \), denoted \( \lambda(k|v_x,v_y) \), depends on intended velocity in orthogonal directions, denoted \( v_x \) and \( v_y \) respectively:

\[
\lambda(k | v_x, v_y) = \exp(\beta_0 + \beta_1 (v_x^2 + v_y^2)^{1/2} \times \cos(\theta - \theta_p))
\]  

(1)

Each neuron has a specific set of values for \( \beta_0, \beta_1 \), and \( \theta_p \). Intuitively, \( \beta_1 \) determines background firing rate, \( \beta_1 \) determines the effect of speed on firing rate, and \( \theta_p \) determines the movement direction that elicits maximum firing rate (the preferred direction). Spike times were generated at 10 ms time steps using the time rescaling theorem [36].
Values for the model parameters were assigned to approximate empirically observed firing rates, direction tuning widths, and speed modulation in primate recordings of MI [7, 27]: $\beta_0 = 2.28$ (unitless), $\beta_1 = 4.67$ s/m. The preferred direction ($\theta_p$) was drawn from the von Mises (circular Gaussian) distribution [1] with mean of 0 degrees and varying $b$ parameter; smaller $b$ parameter corresponds to more uniform distribution in $\theta_p$. Finally, the MM-RSE mapping (as described in Section 1.1.2 and Figure 1A) was used to generate estimates of the intended trajectories from simulated ensemble spiking activity.

To analyze simulated performance, both smoothness (mean of absolute angular velocity) and root mean squared error (RMSE) in velocity were calculated. As an implementation detail regarding the bootstrapping procedure to calculate confidence intervals on smoothness, analyses on experimental data draw with replacement from the collection of smoothness measures computed at every time step, while analyses in simulation draw with replacement from the collection of smoothness measures averaged over every trial (in order to speed the simulation process).

C. Experimental Procedure

In each of two subjects, ensemble spiking activity was obtained from primary motor cortex during center-out reaching movements while grasping a two-link robotic manipulandum [38]. Movements were sampled from the robotic manipulandum at 2.5 ms intervals. The two dimensional workspace included 8 possible targets, uniformly spaced on a 10-cm radius circle ($22.5^\circ$, $67.5^\circ$, $112.5^\circ$, etc.).

A commercial “floating” 16 channel microelectrode array (FMA, MicroProbes for Life Science, Gaithersburg, MD) was used to obtain stable recordings. Action potentials were detected and sorted into single units using threshold crossing and the KlustaKwik technique [39].

Spikes from each neuron were binned at 2.5 ms intervals. Analysis of the tuning curves of these neurons under various force field conditions on the reaching arm was published recently, where additional procedural information is detailed [38]. All animal procedures adhered to National Institutes of Health guidelines on the use of animals and were approved by the MIT Committee for Animal Care.

Our present investigation includes only non-force field trials. Neural activity was modeled using the tuning curve function described in equation (1), for inclusion in calculations of the MM-RSE filter. Parameters of each neuron were fit with training trials using maximum likelihood estimation, implemented by the Matlab function glmfit [27, 40], and decoding was performed on separate test trials.

Data from subject 1 included 12 neurons, with 305 training and 200 test trials. Data from subject 2 included 11 neurons, with 300 training and 200 test trials. Each trial lasted an average of 730 ms (S.D. 150 ms) for Subject 1 and an average of 653 ms (S.D. 103 ms) for Subject 2. Individual trials were segmented into lengths equal to these averages for convenience. For trials shorter than the average, post-target-acquisition movement and neural activity was included. Normalization and interpolation of shorter trials was not pursued – this would have significantly complicated data interpretation. Longer trials were truncated.

III. RESULTS

A. Basic Tuning Curve Properties and Decoding Examples

Tuning curves estimated from experimental data show a spread of tuning width, tuning depth, and firing rates (Figures 2 A-D). Preferred directions are more evenly distributed in subject 1 than in subject 2 where 6 out of 10 neurons orient towards target 4 (Figure 2 C-D). Individual trial decodes (Figure 3) show a spectrum of quality, comparable to a related decoding study from parietal reach region [9].

These single trials show that the RSE approach generates smooth movements, comparable with the true path, whereas the MM-RSE degrades smoothness to the quality of Bayesian decoding with a random walk prior. Also, we see that individual target probabilities dominate to varying extent, generally (but not always) tending to separate towards the completion of the trial (Figures 3 C-D).
While the MM-RSE trajectory appears less smooth than the RSE trajectory in Figure 3A, the difference in smoothness appears less in Figure 3B. Looking at the corresponding posterior probability on targets (Figures 3C, D respectively), we see that the probabilities of true and neighboring targets are smoother and more disparate in Figure 3D compared with Figure 3C, possibly allowing a smoother MM-RSE trajectory in Figure 3B than in Figure 3A in these single trials.

B. Mixture Models Degrade Smoothness

To quantify smoothness, we calculated the average magnitude of angular velocity (Figure 1B), as described in Methods, above. Cumulative density functions (CDF) on the angular velocity (Figure 4) confirm that RSE is significantly smoother than MM-RSE or random walk priors. In subject 2 (Fig. 4B), MM-RSE nearly coincides with the random walk CDF but is significantly less smooth than RSE. Confidence intervals on the various CDF are within the line thickness provided, because each CDF is estimated from a composite of 57,800 time points in Subject 1, and 51,600 time points in Subject 2. These data counts represent 200 reaching trials in each subject. Significance here and elsewhere in the manuscript is asserted on the basis of non-overlapping 95% confidence intervals.

C. Smoothness in Stages of the Filter

Where in the MM-RSE procedure (Figure 1A) does this degradation of smoothness arise? To answer this question, we first tracked angular velocity in a single trial example (Figure 5A), following a single spike occurrence from a single neuron in an otherwise non-spiking ensemble at time step k=131. At k=131, the prediction density of each parallel filter is unresponsive to the spike, because it does not consider neural signals from the current time step. However, the posterior density of each parallel filter does respond immediately, with a small angular velocity. Similarly, the target probabilities, which are based on data likelihoods, adjust to reflect the new spike. Finally, the output filter registers a substantially larger angular velocity at this time step. This single trial suggests that large angular velocities in the output filter do not originate in the response of individual parallel filters to spikes.

To confirm this intuition, we average across multiple trials to compare the angular velocity of individual parallel filters, output filter, and the actual arm movement (Figure 5 B,C). This analysis confirms our intuition, showing that average angular velocity in the output filter significantly exceeds that of individual parallel filters or the actual arm movement. Significance here and elsewhere in the manuscript is asserted on the basis of non-overlapping 95% confidence intervals. A curious, but statistically significant observation in subject 2 is that the parallel filter for target 4 demonstrates slightly smaller angular velocity than other targets (Figure 5C). This likely relates to 6 out of 10 neurons in subject 2 orienting towards this target (Figure 2D). Later on in this paper, we explore this theme using simulation where distributions of preferred direction are varied to demonstrate effect on angular velocity.

D. Relationship between Angular Velocity and Change in Dominant Target Probability

Given the dramatic difference between smoothness in the parallel filters versus output filter, it is unlikely that MM-RSE performance degradation represents a property of the point process filtering component of the overall mapping. An alternate explanation is that abrupt changes in the target probability ripple into the output filter through the weighted averaging step (Figure 1A). We now verify this through mathematical analysis, with subsequent validation on experimental data. To maintain generality, this analysis uses the abstract description of MM-RSE in Figure 1A, without relying on implementation-specific equations within particular examples of MM-RSE such as [2, 3].

Angular velocity \( \frac{\partial \theta}{\partial t} \) can be written as:

\[
\frac{\partial \theta}{\partial t} = \frac{\partial \theta}{\partial \alpha} \frac{\partial \alpha}{\partial t}
\]  

where \( \alpha \) is the probability of the dominant (most probable) target. The \( \frac{\partial \theta}{\partial \alpha} \) term represents the sensitivity of angular
velocity to changes in dominant target probability. The \( \frac{\partial \alpha}{\partial t} \) term represents the change in dominant target probability over time. In the following derivation, we will show that \( \frac{\partial \theta}{\partial t} = \theta_0 \), where \( \theta \) is the angle between output velocity vectors from the parallel filters corresponding to the two most probable targets. We explicitly validate this theoretical result using empirical data (Figure 6).

Let \( u, v \in \mathbb{R}^2 \) be the parallel filter output 2D estimated velocity vectors for the two dominant (two most probable) targets, where \( v \) has target probability \( \alpha \). The filter output 2D estimated velocity vector \( w \in \mathbb{R}^2 \) is an average of estimated velocity vectors for all targets (Figure 1, Step 4), weighted by the corresponding target probabilities. Consequently, the weighted average of \( u \) and \( v \) provides an approximation for \( w \): \( w \approx \alpha v + (1 - \alpha) u \) \( \text{(3)} \)

The error in this approximation is the contribution of the parallel filters corresponding to the remaining targets. Rather than specifically relating to the total number of available targets, this error diminishes with decreasing probability of the remaining (non-dominant) targets.

Let \( \theta_0 \) be the angle between \( u \) and \( v \). Let \( \theta \) be the angle between \( u \) and \( w \). The angle \( \theta \) increases from 0 to \( \theta_0 \) as \( \alpha \) increases from 0 to 1.

We are interested in finding an equation for \( \frac{\partial \theta}{\partial t} \) which is the rate of change of \( \theta \) between \( u \) and \( w \). Note that angular velocity measure depicted in Figure 1B and plotted in Figure 6 is \( \frac{\partial \theta}{\partial t} \) divided by the time step, which is 2.5 ms in our experimental data. To begin, note that the dot product and cross product relate \( \theta \) to \( ||w|| \) and \( |w| \) :

\[
|u||w| \cos \theta = u \cdot w = u^T [\alpha v + (1 - \alpha) u] = \alpha |u \times v|
\]

\( \text{(4)} \)

Divide equation (5) by equation (4):

\[
\tan \theta = \frac{\alpha |u \times v|}{(uv) + (1 - \alpha) |u|^2}
\]

\( \text{(6)} \)

Take the derivative with respect to \( \alpha \).

\[
\sec^2 \theta \frac{\partial \theta}{\partial \alpha} = \frac{|u \times v| [\alpha (uv) + (1 - \alpha) |u|^2]}{[(uv) + (1 - \alpha) |u|^2]^2} - \frac{\alpha |u \times v|(uv - |u|^2)}{[(uv) + (1 - \alpha) |u|^2]^2}
\]

\( \text{(7)} \)

Substitute \( \sec^2 \theta = 1 + \tan^2 \theta \) and equation (6) for \( \tan \theta \) into the left side of equation (8):

\[
\left[1 + \left( \frac{\alpha |u \times v|}{(uv) + (1 - \alpha) |u|^2} \right)^2 \right] \frac{\partial \theta}{\partial \alpha}
\]

\( \text{(9)} \)

\[
\frac{|u \times v||u|^2}{[(uv) + (1 - \alpha) |u|^2]^2}
\]

\( \text{(10)} \)

\[
\frac{\partial \theta}{\partial \alpha} = \frac{|u \times v||u|^2}{[(uv) + (1 - \alpha) |u|^2]^2} + \frac{2 \alpha |u \times v|^2}{[(uv) + (1 - \alpha) |u|^2]^2}
\]

\( \text{(11)} \)

\[
\frac{\partial \theta}{\partial \alpha} = \frac{2 |u||v| \sin \theta_0}{\alpha^2 |v|^2 + 2 \alpha (1 - \alpha) |u||v| \cos \theta_0 + (1 - \alpha)^2 |u|^2}
\]

\( \text{(12)} \)

\[
\frac{\partial \theta}{\partial \alpha} = \frac{|u||v| \sin \theta_0}{\alpha^2 |v|^2 + 2 \alpha (1 - \alpha) |u||v| \cos \theta_0 + (1 - \alpha)^2 |u|^2}
\]

\( \text{(13)} \)

Figure 6. Changes in dominant target probability as a determinant of angular velocity. The approximately linear relationship between these two variables is predicted by equation (14). The predicted slope (green line) is in close agreement with the empirical slope (red line). Trends are replicated across (A) subject 1 and (B) subject 2. In (A) the empirical slope = 26 deg/ms/unit probability with y-intercept 0.08 deg/ms. The predicted slope = 18 deg/ms/unit probability with y-intercept 0 deg/ms. L1 error = 0.09 with 95% CI=[0.01, 0.43] (deg/ms), and \( R^2 = 0.70 \). In (B) the empirical slope = 23 deg/ms/unit probability with y-intercept 0.04 deg/ms. The predicted slope = 18 deg/ms/unit probability with y-intercept 0 deg/ms. L1 error = 0.07 with 95% CI=[0.00, 0.37] (deg/ms), and \( R^2 = 0.74 \).
Assume that parallel filter output speeds are comparable, \( |u| \approx |v| \). Furthermore, assume that \( \theta_0 \) is sufficiently small that \( \sin \theta_0 \approx \theta_0 \) and \( \cos \theta_0 \approx 1 \). The small angle assumption is not directly dependent on the number of targets. Because each of \( u \) and \( v \) tend to point in the direction of the displacement vector between current position and the corresponding target, this small \( \theta_0 \) assumption is expected to break down when the corresponding displacement vectors are poorly aligned.

In target sets with distinct target locations, the small angle assumption tends to be violated towards the end of a reach. In the case where displacement vectors are markedly divergent at the beginning of the reach, such as with two diametrically opposed target positions, target probabilities more easily separate earlier in the reach, resulting in a single dominant target. Based on this final set of assumptions that \( |u| \approx |v| \) and \( \theta_0 \) is small, Equation (13) simplifies to:

\[
\frac{\partial \theta}{\partial \alpha} \approx \theta_0
\]  

Equation (14) together with equation (2) predict that heading direction will increase linearly with change in dominant target probability at a rate of \( \theta_0 \).

In Figure 6, we explicitly validate this theoretical prediction using motor cortex spiking neural data recorded from two separate subjects during reaching movements. We regress the angular velocity against the first difference in probability of the dominant target, demonstrating a reasonable approximation by various error measures (L1 norm, \( R^2 \)). The slope also gives us some sense for the strength by which change in target probability influences angular velocity in the output filter. Specifically, a 10% change in probability of the dominant target results in appreciably large output filter angular velocities of 23 and 26 degrees per millisecond for subjects 1 and 2 respectively.

The slope of the Figure 6 regression reasonably matches the prediction made by equation (14). Because eight reach targets are separated evenly, \( \theta_0 = 45^\circ \) for large parts of the reach. The theory-predicted slope is \( 45^\circ / 2.5 ms = 18^\circ / ms / \text{unit change in probability of } \alpha \) because the x-axis plots the first difference in dominant target probability, which is equal to \( \Delta \times \frac{\partial \alpha}{\partial t} \) where \( \Delta = 2.5 ms \). For comparison, the empirically-derived slopes are 26 and 23 degrees per millisecond per unit change in probability of dominant target in subjects 1 and 2 respectively.

E. Other Determinants of Angular Velocity: Target Probability and Target Spacing

Do smoother trajectories result from greater certainty in target? The theoretical answer to this question is not simple, relating to both the \( \frac{\partial \theta}{\partial \alpha} \) and \( \frac{\partial \alpha}{\partial t} \) terms that determine trajectory smoothness \( \frac{\partial \theta}{\partial t} \) in equation (2). It may be the case that larger \( \alpha \) result in smaller \( \frac{\partial \alpha}{\partial t} \) by making the target probability computation less sensitive to incoming data (Figure 1A, Step 3). However, the second term in equation (2) remains problematic. This is because, in general, \( \frac{\partial \theta}{\partial \alpha} \) is not always smallest when \( \alpha \) is close 1.
To see this, let $\alpha_0$ be the value (not necessarily confined to $0 < \alpha_0 \leq 1$) in equation (3), i.e.

$$w \approx \alpha v + (1 - \alpha) u = u + \alpha (v - u)$$

for which $w$ is perpendicular to $v - u$. $\frac{\partial \theta}{\partial \alpha}$ is largest at $\alpha = \alpha_0$ and decreases monotonically as $\alpha$ moves away from $\alpha_0$ (see proof in Appendix).

In the `well-behaved" case where $\alpha_0 \in [0,1)$ it is the case that $\alpha = 0$ and $\alpha = 1$ are local minima of $\frac{\partial \theta}{\partial \alpha}$ on the domain $[0,1]$, and one of them is the absolute minimum. In the `not-well-behaved" case, $\alpha_0 \not\in [0,1)$, one of the endpoints ($\alpha = 0$ or $\alpha = 1$) is the absolute minimum and the other is the absolute maximum. As such, it is not the case that $\alpha = 1$ will always be the best option for reducing $\frac{\partial \theta}{\partial \alpha}$.

On the other hand, it is plausible that smoother trajectories result from more narrowly spaced targets. In section 3.4 above, we showed that $\frac{\partial \theta}{\partial \alpha} \approx \theta_0$. Because the output estimated velocities of the parallel filters generally point toward their corresponding targets, it follows that decreasing the spacing between targets would decrease $\theta_0$, reducing both $\frac{\partial \theta}{\partial \alpha}$ and $\frac{\partial \theta}{\partial t}$, resulting in smoother targets. We have anecdotaly seen improved smoothness by completely aligning target locations (targets placed at the same spatial location, but at different points in time), as with our previous work that applies mixture models to reaching movements with variable arrival time [13].

**F. Smoothness, Neuron Count and Tuning Curve Distribution**

While section 3.5 focused on $\frac{\partial \theta}{\partial \alpha}$, equation (2) implies that abrupt changes in target probability ($\frac{\partial \alpha}{\partial t}$) can also contribute to MM-RSE performance degradation. The quantity $\frac{\partial \alpha}{\partial t}$ in turn is determined by the neural data. For spiking data, low channel count and unevenly spaced preferred directions contribute to abrupt changes in the ensemble data likelihood, driving $\frac{\partial \alpha}{\partial t}$. In Figure 7, we examine the relative contributions of neuron count and tuning curve distribution to output filter smoothness in simulation as a preliminary step towards improving MM-RSE performance. These curves suggest that output filter smoothness is predominantly improved by increasing neuron count, and comparatively less improved by employing more uniform tuning curve distributions (Figure 7C). We verify the robustness of these results by replicating these graphs at 3 orders of magnitude in increment covariance (Figure 8), where increment refers to the additive Gaussian random variable in a linear Gaussian state equation [7]. These trends are generally preserved, although ensemble size has minimal effect on angular velocity at increment covariance of 10 cm²/s², likely due in large part to the significant overall deterioration in performance related to tracking faster movements.

Despite the marginal effect of $b$ parameter on trajectory smoothness, increasing the uniformity of tuning curve distribution is likely still relevant to other performance measures, as suggested by its substantial effect on open-loop root mean squared error (RMSE) in velocity (Figure 7B). Differences in open-loop RMSE between biased and unbiased estimators (such as the comparison between population vector analysis (PVA) and optimal linear estimation) may not be relevant to closed loop performance [33]. However, increased open-loop RMSE among unbiased methods (such as our recursive Bayesian filters) will likely degrade closed-loop performance [23, 35]. Unlike the systematic decoding error of PVA bias, decoding errors represented by RMSE in unbiased estimators are unpredictable, and likely difficult for the user to compensate through feedback.
IV. DISCUSSION

The use of mixture models is a principled strategy in the design of brain-machine interface (BMI) algorithms. In this paper, we demonstrated that mixture models actually degrade trajectory smoothness when combined with reach state equations (RSE) to support unknown targets from a discrete target set. We focused on angular velocity as a meaningful and convenient measure of smoothness. We combined empirical and mathematical analysis to identify abrupt changes in target probability as a root cause of MM-RSE performance degradation. Our results demonstrated that angular velocity in the trajectory is approximately proportional to change in target probability. The constant of proportionality equals the difference in heading between parallel filters corresponding to the two most probable targets, suggesting a smoothness benefit to more narrowly spaced dominant targets. In contrast, smoothness is not directly related to number of targets, although a separate consideration in scaling MM-RSE to large numbers of targets is a linear increase in parallel filters (Figure 1, Step 2). Decoding movements to a continuum of target locations with a single Gaussian prior was addressed by the generative reach state equations, which does not require mixture modeling [2, 7, 10-14].

We used simulation with a point process model of motor cortical spiking to explore beyond the conditions tested in the experimental data. These simulated results suggested that increasing neuron number and uniformity of tuning curve distribution can mitigate this degradation to varying extent. Increasing neuron count benefited both trajectory smoothness and RMSE in velocity. While uniformity of tuning curve distribution only marginally improved trajectory smoothness, it substantially modulated RMSE in velocity. An alternate approach that was not investigated here is to attempt to maximize certainty in the target identity through side information such as gaze analysis coupled with computer vision [2], or planning regions such as dorsal premotor cortex [5] and parietal reach region [41].

In practice, neuron number or channel count (with electroencephalogram (EEG), electrocorticoencephalogram (ECoG), or local field potentials (LFP)) is not easily increased because it requires the fabrication and implantation of more electrodes. Similarly, it might seem that tuning curve distribution is a fixed quantity. However, recent work has shown that tuning curves can shift in response to closed-loop training with fixed BMI decoding parameters [19, 42]. This represents a prime opportunity to enforce an optimal distribution of BMI decoder tuning curves, allowing the user to train to conform to this distribution in order to achieve optimal performance. This training process might involve a sequence of adjustments in the BMI decoder tuning curves in order to allow the user to progressively rotate their neuronal tuning curves towards the optimal distribution. In this proposed approach, neuronal properties are trained towards parameter regions that result in optimal open-loop decoding in the fully-learned state, in an effort to minimize the need for excessive closed-loop corrective behavior. More immediately, a subset of the full ensemble might be determined in part based on the uniformity criterion.

V. APPENDIX

Here, we justify the statement from Section 3.5 that
\[ \arg \max_a \frac{\partial \theta}{\partial \alpha} = \alpha_0. \] The parallel filter velocity estimates \( u \) and \( v \), and approximate weighted output \( w \) are identically defined in section 3.5. Recall \( w = (1 - \alpha)u + \alpha v = u + \alpha(v - u) \). Define \( p = u + \alpha_0(v - u) \), where \( \| p \| = r \) is the perpendicular distance from the origin to the line passing through points given by \( u \) and \( v \). Define \( a \) to be the signed difference from \( p \) to \( u \) with magnitude \( \| u - p \| \). Define \( x \) to be the signed difference between \( w \) and \( p \) with magnitude \( \| w - p \| \). One configuration of these variables is depicted in Figure 9, ordered from left to right as \( (u, p, w, v) \). The remaining canonical configurations of Figure 9 correspond to the orderings \( (p, u, w, v) \), \( (u, w, p, v) \), and \( (u, w, v, p) \). The sign of \( a \) is defined to be negative if \( u \) appears before \( p \) in the ordering, and vice versa. The sign of \( x \) is defined to be positive if \( p \) appears before \( w \) in the ordering, and vice versa. Using the signed differences \( a \) and \( x \), the relation \( x \equiv \| w - u \| + a \) holds for all four configurations.

To show \( \arg \max_a \frac{\partial \theta}{\partial \alpha} = \alpha_0 \), first calculate \( \frac{d \theta}{d \alpha} \). Note that
\[ \tan \theta = \frac{x}{r}. \] Now substitute \( x = \alpha \| v - u \| + a \). This equality follows from \( \| w - u \| = x - a \) and \( w = u + \alpha(v - u) \). Combining these equalities, \( \tan \theta = \frac{\alpha \| v - u \| + a}{r} \). Differentiating both sides...
with respect to $\alpha$ gives $(\sec^2 \theta) \frac{d\theta}{d\alpha} = \frac{\|v-u\|}{r}$. Rearranging terms, $\frac{d\theta}{d\alpha} = \frac{\|v-u\|}{r} (\cos^2 \theta)$. Thus $\frac{d\theta}{d\alpha}$ is maximized by $\theta = 0 \leftrightarrow v = p \leftrightarrow \alpha_0 \leftrightarrow \arg \max_\alpha \frac{d\theta}{d\alpha} = \alpha_0$.

VI. ACKNOWLEDGEMENT

A.W. was supported by the Caltech Summer Undergraduate Research Fellowship. K.K. was supported by the UCLA Amgen Scholars Program. L.S. was supported by the American Heart Association Scientist Development Grant (11SDG7550015), DARPA Reliable Neural-Interface Technology (RE-NET) Program, and the UCLA Radiology Exploratory Development Grant.

VII. REFERENCES


Siamak Yousefi received his Ph.D. from University of Texas at Dallas. He is currently a postdoctoral fellow at the Hamilton Glaucoma Center, University of California at San Diego where he develops ophthalmic image and data analysis methods. His research areas include biomedical image analysis, pattern recognition, and machine learning. He is a member of the IEEE and ARVO.

Alex Wein received the B.S. degree with honors in Computer Science and Mathematics from the California Institute of Technology in 2013. He is presently a PhD candidate in Mathematics at MIT. He contributed to this work during his undergraduate studies while working as a researcher at the Neural Signal Processing Laboratory, Department of Radiology, University of California, Los Angeles.

Kevin Kowalski received the B.S. degree with honors in Computer Science from the California Institute of Technology in 2012. He is currently a PhD candidate in Computer Sciences at the University of Wisconsin-Madison. He contributed to this work during his undergraduate studies while working as a researcher at the Neural Signal Processing Laboratory, Department of Radiology, University of California, Los Angeles.

Andrew Richardson is co-director of the Translational Neuromodulation Laboratory, Department of Neurosurgery, University of Pennsylvania. He received the B.S.E. degree in biomedical engineering from Case Western Reserve University, Cleveland, OH, in 2000, and the S.M. degree in mechanical engineering and the Ph.D. degree in biomedical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 2003 and 2007, respectively. He was a Postdoctoral Associate at MIT in 2007–2008, and a Senior Fellow at the University of Washington Seattle in 2008-2012. His research interests include motor learning, neuroplasticity, and neuromodulation.

Lakshminarayan Srinivasan is the principal investigator at the Neural Signal Processing Laboratory, Department of Radiology, University of California, Los Angeles. He received the B.S. degree with honors in Electrical and Computer Engineering from the California Institute of Technology in 2002. He obtained the S.M. and Ph.D. degrees in Electrical Engineering and Computer Science from MIT in 2003 and 2006 respectively. He received the M.D. from Harvard Medical School in 2009, obtained physician licensure from the Medical Board of California in 2010, and board eligibility from the American Board of Radiology in 2014.

From 2006 to 2009, Dr. Srinivasan was a postdoctoral research fellow with the Department of Neurosurgery, Massachusetts General Hospital, Boston, MA. During 2009-2010, he was a Visiting Scientist with the Computational Neuroscience Laboratory at the Salk Institute, La Jolla, CA. In 2010, he started the Neural Signal Processing Laboratory, which seeks to advance the scientific and therapeutic understanding of the human brain by combining information science, neuroscience, and medicine.