nature portfolio

Peer Review File

Reach-dependent reorientation of rotational dynamics in motor cortex

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The manuscript of Sabatini and Kaufman reports a data analysis method inspired by a dynamical systems perspective of neural dynamics in (pre)motor cortex during short transient reach movements. Different to previous methods of similar kind, the current approach respects differences between different reach conditions instead of projecting the high-dimensional data onto a single low-dimensional manifold. The authors demonstrate that by allowing separate low-dimensional manifolds ("planes") for each reach conditions they can explain much more variance in the data and accordingly achieve higher decoding performance when decoding hand position based on neural activity. They also show that within the planes for each reach condition there is still evidence for rotational dynamics during reach, as suggested previously for the previous "one-fits-it-all" approach.

The manuscript is very well written such that the rational of the study and the approach are very clear. Also the methods look very solid and the figures are very illustrative (sometimes almost to the point that it gets hard to find the actual empirical data between all the hypothetical drawings for illstrating the method - please reconsider this style). If this was an original study, I would probably be enthusiastic about it, since I like the way it makes complex data intuitively accessible and quantitatively describable. What needs attention, though, is the fact that the data has been published at least three times already (acknowledged in the manuscript). The important question is what we actually learn about motor cortex that was not known before from these or other previous studies. I find it remarkable that the only time that the manuscript phrases a hypothesis, it is a hypothesis about why the previous methods applied to the same data performed so poorly. If I wanted to be sarcastic, I would have to wonder: if the previous poor description of the data by the same authors would not have been published in a highest profile journal, what would be the research questions that defines the relevenace of the current study? More seriously: I do believe that the current approach is worth presenting, but I am not convinced of the current framing, basically motivating it by a previously applied mcuh poorer approach of the same authors, and I am not sure if in its current form the selected journal is the right place for it. Maybe applying the approach to different data sets to demonstrate its general validity might be a way out (see below).

What the new analysis reveals is that the previous publications of the same data in a dramatic way glossed over important detail, with the result of a clean and simple (but almost misleading) story. By revisting the data, the authors now reveal the actually much higher complexity of the data that many readers and scientists who preferred more classical approaches always tried to emphasize. Since the previous jPCA approach only explained about 1/10 of the variance in the data, it was clear from the beginning that it cannot be the final truth and still it was used to claim that motor cortex would undergo simple rotational dynamics during a fast, transient reach, supporting the validity of the dynamical systems view. The new study tries to "rescue" the core of the idea. While it admits the much higher variability induced in the neural states between different reach conditions (acknowledging the fact that there might indeed be some form of "representational" information about the different reaches), by identifying different manifolds for each condition, it still emphasizes the rotational dynamcis within these manifolds.

The former observation (condidition-dependent initial conditions) has been suggested and shown before (Perich et al. 2020; Michaels et al. 2020; acknowledged in the manuscript). So the important questions is: is the latter (rotational dynamics within the condition-specific planes) a non-trivial finding, what would be the alternative hypothesis, and which functional aspect of motor cortex do we understand better by this form of description? Since we are looking at a behavior that starts with a resting arm and ends with a resting arm, while briefly producing a pattern of muscle activations for reach in between, one wonders what the alternative is to neural dynamics that more or less (i.e. except for pose-dependend modulations) return to their initial conditions after running along a brief trajectory in neural state space? At the individual neuron level in motor cortex it is a well-known fact that many neurons start with no/low level of activity prior to reach and end with this level after the reach, showing transient, often biphasis activation during acceleration/decelaration in between. One non-trivial finding probably is the observed conservation of eigen-frequencies in the rotational dynamics, since the authors say (page 3): "kinematic parameters performed poorly as predictors of eigenvalues". Unfortunately, it is not clear to me what the basis for this statement is (which analysis of the reach velocity profiles across different conditions shows this?) and what the functional interpretation/biological relevance of these preserved eigenvalues is?

The higher levels of explained variance per se, compared to other linear methods, is impressive, but I consider this an incremental improvement unless the here presented model explains something about motor cortex function that could not be unvealed with other methods, including non-linear methods. For example, in Figure 7, the encoding and decoding methods are not compared to other high-performing algorithms within the same domain modelling dynamical systems without the assumption of shared rotational frequencies, such as autoLFADS (Keshtkaran et al. 2022). Also, I think the method presented here would have to be demonstrated to generalize well to motor cortex data during different type of motor behavior.

In Figure 8, the author asserts that their LDR method surpasses standard decoding by predicting activity in both the output-potent and output-null dimensions. However, none of the preceding figures differentiate between the potent subspace and null subspace, making the conceptual explanation disconnected from the remaining findings presented in the manuscript.

Keshtkaran MR, Sedler AR, Chowdhury RH, Tandon R, Basrai D, Nguyen SL, Sohn H, Jazayeri M, Miller LE, Pandarinath C. A large-scale neural network training framework for generalized estimation of single-trial population dynamics. Nat Methods. 2022; 19(12):1572-1577.

Michaels, J. A., Schaffelhofer, S., Agudelo-Toro, A. & Scherberger, H. A goal-driven modular neural network predicts parietofrontal neural dynamics during grasping. Proc Natl Acad Sci U S A 117, 32124–32135 (2020).

Perich, M. G. et al. Motor cortical dynamics are shaped by multiple distinct subspaces during naturalistic behavior. http://biorxiv.org/lookup/doi/10.1101/2020.07.30.228767 (2020) doi:10.1101/2020.07.30.228767.

Reviewer #2 (Remarks to the Author):

This is an interesting article that presents an ensemble of ambitious analyses of motor cortical activity during a large variety of reaches.

These extensive analyses shed new light on several aspects of motor cortical dynamics.

Specifically, the authors uncover that diverse reach conditions are associated with more diverse dynamical features - notably, oscillation planes and locations in neural state-space - than reported in previous reach studies. In addition, the authors discuss possible consequences for the expressivity and generalization abilities of motor cortical computations, which are of broad interest.

The article is already a valuable read, and I think that it can be further refined in a relatively straightforward manner by adding a few clarifications about the methodology used and its implications on how results are interpreted.

Main comments:

1. One main notion introduced in the manuscript is that of 'condition', as the authors compare various quantities across and within conditions. However, occasionally, authors also plot quantities colored by 'target angles' (for instance, for the eigenvalue analysis in extended data fig. 2).

After reading the manuscript, I still had some uncertainty about how 'conditions' and 'target angles' relate. The method says 'On some trials, the monkey was required to avoid virtual barriers presented at the same time as the target, eliciting curved reaches to produce 72 different reaching “conditions” (36 straight, 36 curved).'. Is one 'condition' several target angles but a single type of barrier and/or extent required? Or does a 'condition' refer to the combination of a single target angle, and a type of obstacle / reach extent required?

2. The authors state that 'As previously shown, though, these rotational dynamics explained only 7-12% (s.d. < 13% across conditions; Extended Data Fig. 1) of the variance in peri-movement firing rates', and that this relates to a single LDS shared across conditions.

However, Lara, Cunningham and Churchland (Nature Communications, 2018) report that - across different reach angles - a single linear dynamical system fits the *reach-angle-dependent* part of M1 activity with an r-squared of R2 = 0.84 and 0.76 (and R2 = 0.74 and 0.62 when constraining the dynamical matrix to be skew-symmetric). I don't think that there is a direct comparison for this specific analysis in your case (you come close when using jPCA across conditions, but it is still a different analysis). Of course, your subsequent observation of condition-specific oscillation planes would suggest that a dynamical fit across conditions would be poor in your data. If that's true (and it'd be nice to directly test this, even if using a simpler analysis than in Lara et al. and just fitting a linear dynamical system across conditions), is it because you are considering a larger variety of conditions than in the Lara paper? Indeed, in the current paper you have obstacles, while Lara et al. didn't (the different reach angles are what they consider as 'conditions' during the movement period). Is the extent to which conditions differ in dynamics related to how much they differ in terms of kinematics and/or (inferred) EMGs? This could be a step towards providing evidence for the idea mentioned in the discussion that 'allowing the population state to rotate in substantially different planes for different reaches may allow motor cortex to produce a variety of different output signals across reaches'.

3. The authors perform several advanced analyses to try to estimate quantities that can be related to models of motor cortical computations. This is a worthwhile but difficult undertaking, notably because the statistical methods used for estimation have unknown bias and variances, that could differ across the quantities of interest. To mitigate these issues, as a general guideline, it might be best to use statistical tools that are as directly related as possible to the question that one is asking. Notably, there are tools that can fit latent linear dynamical systems directly from PSTHs, an approach that also allows computing confidence intervals and cross-validation across different trials, without relying on a Poisson assumption - see for instance O'Shea, Duncker et al., bioRxiv 2022 (Method 9. Latent linear dynamical system model). In contrast, here, it appears that the authors first use a dimensionality reduction method (paragraph 'Fitting dynamics to a single condition' of the methods) prior to fitting the dynamics. I suspect that the two methods may give slightly different results. Specifically, given that with the former method dimensionality reduction is not done a priori but is rather a consequence of the rank of the fitted connectivity matrix, this method might be able to extract more precise estimates of eigenvalues and eigenvectors. It'd be great to at least discuss how the method chosen to estimate the quantities of interest may affect the results.

4. Relatedly, it may be difficult to conclude whether eigenvalues are significantly different across conditions - indeed, the estimated magnitude of eigenvalue distances due to noise (fig. 3c, noise) is similar to the magnitude of the eigenvalue distances across the eigenmodes estimated from single-condition data (fig. 3a and h).

Comparing distances between pairs of eigenvalues introduces the difficulty of knowing which eigenvalue corresponds to which, and of relying on a Poisson noise assumption - which may contribute to the above observation.

As an alternative to comparing distances between pairs of eigenvalues, it could be interesting to compare, between conditions, the estimated eigenvalue density quantifying the uncertainty in eigenvalue position due to trial sampling (as in O'Shea, Duncker et al., bioRxiv 2022). For instance, one could use a Kolmogorov–Smirnov test on the projection of eigenvalue densities along either the imaginary or real axis.

A similar approach could also be used for eigenvector angles, where the advantage would be to get rid of the Poisson assumption and instead directly estimate the uncertainty due to sampling.

5. Two parts of the discussion were unclear to me.

First, you mention that 'Smoothly varying the rotational plane may allow motor cortex to generate the “correct” activity patterns for new reaches by reaping the benefit of strong generalization due to local linearity'. On the one hand, in a (quasi)-linear regime, previous work (e.g. Lara et al., Nature Communications 2018; Vyas et al., Neuron 2018; Vyas et al., Neuron 2020; Logiaco et al., Cell Reports 2021) has emphasized smoothness and/or generalization properties of the dynamics when simply changing the initialization of the circuit to create movement variants. If this is what you are referring to, it would be nice to clarify. If it is not, it would be nice to clarify this alternative idea further because the cited references either do not refer to dynamical cases or to quasi-linear dynamics, so I do not understand what the authors mean.

Second, I am also confused about the discussion around corticospinal neurons (CSN). First, you could have several separate subnetworks with linear dynamics (possibly, each of them having rotational dynamics) that each project to a separate CSN or CSN population. This mechanism could separately modulate the phase and amplitude of each CSN, which seems to contradict your statement that 'Readouts from rotational dynamics are strictly locked in phase and magnitude, meaning that spiking activity of multiple CSNs cannot be independently modulated'. Second, given that your results concern a nonlinearity *across conditions, not across time* while within each condition the activity is well-described by linear dynamics, I do not understand how the nonlinearity could help control different CSNs that are simultaneously active during one condition.

More generally, it could be good to discuss that one difficulty in linking the authors' observations to interpretations about M1's function is that the output of M1 is not directly experimentally accessible (we know it is a signal sent to the muscles, but this signal is only indirectly related to the recorded kinematics).

Additional remarks:

1. Is the fact that you are able to better decode kinematics from the LDR related to previous work that has shown how BMI decoding can benefit from accounting for the dynamical nature of M1 activity (e.g. Kao, Ryu and Shenoy, Scientific Reports 2017; Kao, Nuyujukian, Ryu, Churchland, Cunningham and Shenoy; Nature communications 2015)?

2. The discussion refers to 'autonomous' dynamics, but I am not sure why the argument matters here. If the authors mention it, it would be nice to clarify that there could really be many meanings behind this term. Notably, are the authors referring to the hypothesis that M1 is mostly driven by its own recurrent connections (as opposed to strongly interacting with other brain regions), or to the notion of feedforward control from the optimal control literature which differentiates whether or not the controller receives fast and reliable sensory feedback from the external world?

3. You are mentioning the danger of 'overfitting' when fitting dynamics. As mentioned above, there are ways to separate training and test sets to control for this possibility.

4. For Fig. 3h, you mention that 'The projections from each condition’s motor cortex activity recovered the temporal basis functions almost perfectly (91-95% variance explained)'. Does that refer to the mean projection across conditions? The variance around the individual translucent traces in Fig. 3h seems large, so I would expect the error computed per condition to be larger.

Reviewer #3 (Remarks to the Author):

Review of: Reach-dependent reorientation of rotational dynamics in motor cortex

The manuscript presented by Drs. Sabatini and Kaufman builds upon previous work describing rotational dynamics in primate motor cortical single unit ensembles. The critical innovation is expanding the model of rotational dynamics to examine multiple planes and frequencies. While frequencies remained relatively stable across conditions, the planes of rotation varied more than would be expected by chance. Modeling rotations in multiple planes (location dependent rotations, LDR) greatly increases the explanatory power of the resulting dynamical system models. Models with larger numbers of parameters are expected to explain greater degrees of variance (as the authors correctly point out in the results section). However, the increase in explained variance from < 12% to > 90% suggests that the new model is capturing important sources of variance ignored by previous formulations. Furthermore, standard models applied to individual conditions (single planes) also explain a high fraction of condition-specific variance. The results presented suggest that the increase in variance explained is tied specifically to adding additional rotation planes, but not frequencies (which are kept consistent in the final LDR model). I think this point could be made more explicit by making this comparison directly in a figure: I was expecting to find a figure comparing variance explained by taking into account rotations in a single plane (i.e. jPCA), an LDR model fitting the same frequencies in multiple planes (the presented model), and one fitting both planes and frequencies. Figure 2D presents part of this comparison by contrasting jPCA with single condition models, but it feels like it does not cover the full set of comparisons.

The authors introduce the Subspace Excursion Angles (SEA) metric, which orders subspaces according to the angles between them. Using this strategy they identify 6-20 distinct planes with angles > 45 °to each other in motor cortex activity. The authors claim that “Existing methods of estimating dimensionality do not distinguish slight variations from more substantial variations, as long as the occupancy of each additional dimension is above the noise level.” While I believe this is true of standard projection-based methods such as PCA, I am not certain it applies to fractal intrinsic dimensionality estimation methods. While the SEA method seems intuitive and practical, I think it might be useful to relate these results to more standard metrics. For references to possible alternative algorithms, I would suggest looking at:

1. Camastra, Francesco. "Data dimensionality estimation methods: a survey." Pattern recognition 36.12 (2003): 2945-2954.

2. Facco, E., d’Errico, M., Rodriguez, A. et al. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. Sci Rep 7, 12140 (2017).

The relationship between neural activity and movement kinematics is explored using a “sequence to sequence” (StS) encoder framework, i.e. relating single trial dynamics to the set of kinematics (summarized through dimensionality reduction) for a complete reaching motion. This approach can be used to make predictions in both directions. In the final section of the manuscript, the authors highlight the prediction of single trial kinematics from firing rates processed through LDR. The results presented in figure 7F suggest that LDR can greatly improve performance compared to standard instantaneous decoding methods. However, in the final paragraph of the discussion, the authors point out that StS encoders cannot be used for real-time control. This makes the comparison in 7F fall a bit flat. Comparing instantaneous prediction using limited time windows to predictions from one full sequence to another does not seem like a fair comparison, since less information is taken into account for each discrete estimate in the former case. It could be useful to add a comparison with some other kind of StS decoder (perhaps using a deep learning framework). I think this point should be addressed more fully in the discussion. Perhaps the authors could also speculate on ways incorporate StS decoding to BCI control. Maybe a complete reach to grasp action could be decoded (and even displayed in advance using augmented reality) before it is carried out by a robotic limb. Recent work in speech reconstruction could potentially be better suited for an StS approach, given the intrinsic sequential nature of language.

Overall, I think the presented work constitutes an impressive advance in our understanding of motor cortex encoding, presenting an elegant model that greatly expands the explanatory power afforded by previous work. If the three relatively minor points outlined above are addressed, I would fully support publication of the manuscript in Nature Communications.

Carlos Vargas-Irwin

Assistant Professor

Brown University Neuroscience Department

P.S. I also found what appears to be a Typo in the method section:

“We attempted therefore “recovered” the temporal basis functions from neural activity as

(6) B’(c) = L(c)†X(c)

where B’(c) is the recovered temporal basis functions for condition c and † indicates pseudo-inversion.”

We thank the reviewers for their time, for their careful attention, and their helpful comments. The reviewers had two primary concerns. The first concern was that our methods for estimating several important quantities in our data seemed to include arbitrary or sub-optimal choices, or were not sufficiently compared to other methods. On this note, R2 highlighted existing techniques for more efficiently estimating quantities of interest, and R3 suggested several standard methods to compare our methods against. We have now followed these excellent suggestions. The second concern regarded our interpretations of our decoding analysis (R3 and R1), and the relationship to previous literature and alternatives (R1). We have revised our language, clarified our interpretations, and added several new controls to improve interpretability. Together, we think these changes have substantially strengthened our paper.

Reviewer comments appear in blue, our replies appear in black, and quotations from the manuscript appear in red.

Reviewer #1:

The manuscript of Sabatini and Kaufman reports a data analysis method inspired by a dynamical systems perspective of neural dynamics in (pre)motor cortex during short transient reach movements. Different to previous methods of similar kind, the current approach respects differences between different reach conditions instead of projecting the high-dimensional data onto a single low-dimensional manifold. The authors demonstrate that by allowing separate low-dimensional manifolds ("planes") for each reach conditions they can explain much more variance in the data and accordingly achieve higher decoding performance when decoding hand position based on neural activity. They also show that within the planes for each reach condition there is still evidence for rotational dynamics during reach, as suggested previously for the previous "one-fits-it-all" approach.

The manuscript is very well written such that the rational of the study and the approach are very clear. Also the methods look very solid and the figures are very illustrative (sometimes almost to the point that it gets hard to find the actual empirical data between all the hypothetical drawings for illstrating the method - please reconsider this style).

We thank the reviewer for their kind words. Regarding our figures, we have added labels to figures to highlight which sub-panels are conceptual illustrations, and which are data.

If this was an original study, I would probably be enthusiastic about it, since I like the way it makes complex data intuitively accessible and quantitatively describable. What needs attention, though, is the fact that the data has been published at least three times already (acknowledged in the manuscript). The important question is what we actually learn about motor cortex that was not known before from these or other previous studies. I find it remarkable that the only time that the manuscript phrases a hypothesis, it is a hypothesis about why the previous methods applied to the same data performed so poorly. If I wanted to be sarcastic, I would have to wonder: if the previous poor description of the data by the same authors would not have been published in a highest profile journal, what would be the research questions that defines the relevenace of the current study? More seriously: I do believe that the current approach is worth presenting, but I am not convinced of the current framing, basically motivating it by a previously applied mcuh poorer approach of the same authors, and I am not sure if in its current form the selected journal is the right place for it. Maybe applying the approach to different data sets to demonstrate its general validity might be a way out (see below).

The reviewer is of course correct that these data are not new, and that we are indeed continuing in the dynamical systems approach most notably taken by Churchland 2012. While rotational dynamics have been enormously influential in the field for over a decade, and therefore revising this view is itself important, we argue that we have learned three additional important new things from our findings. First, we have better learned the nature of the dynamical system. Specifically, location-dependent rotational dynamics is quite different from any dynamics that have been considered anywhere else to our knowledge, and fits the data much better than ordinary linear dynamical systems. This, by itself, we think is important. However, this finding leads to several other critical insights. This class of dynamics allows for much richer outputs than Churchland 2012-style planar rotational dynamics (see Figure 8), which enables the system to produce a wider possible repertoire of muscle outputs and thereby enables more flexible control. Finally, it links representation to dynamics: the representation sets the location in state space, then the dynamics apply and produce the time-varying coordinated command signals. This helps us understand how the brain solves the "inverse problem" – turning a representation of the desired movement into the needed command signals. This insight also therefore links the representational and dynamical systems approaches: it describes how a fed-forward representation can set the system up to produce the complex, coordinated, time-varying outputs needed to drive the movement itself.

We now make this novelty clearer:

Abstract:

> Our "location-dependent rotations" model fits nearly all motor cortex activity during reaching, and high-quality decoding of reach kinematics reveals a hidden linear relationship with spiking. Varying rotational planes allows motor cortex to simply produce richer outputs than possible under previous models. Finally, our model links representational and dynamical ideas: representation is present in the state space location, which dynamics then convert into time-varying command signals.

Introduction:

> These findings enable several important advances: they allow us to account for virtually all neural variance in dorsal premotor (PMd) and primary motor cortex (M1) during reaching; describe a new class of dynamics that reconciles previously-conflicting interpretations of motor cortex; enable high-fidelity encoding and decoding between motor cortex and kinematics with linear methods; and improve on rotational dynamics as previously understood by allowing for a much richer repertoire of motor cortical outputs.

Discussion: reworked throughout.

Regarding phrasing our findings in terms of hypotheses, we experimented heavily with language when writing this paper. It was very challenging to make this story clear enough for a broad audience, much of which can't be expected to follow the equations. Although we did test many hypotheses in this work, for clarity, we generally chose to jump straight to the conclusion. We found this approach made it much easier to follow the narrative than presenting the possibilities on an equal footing before showing which was supported by the evidence. But, the reviewer makes a point that many readers are likely to miss this, and we have now framed alternative models and hypotheses as such where we could without compromising clarity.

What the new analysis reveals is that the previous publications of the same data in a dramatic way glossed over important detail, with the result of a clean and simple (but almost misleading) story. By revisting the data, the authors now reveal the actually much higher complexity of the data that many readers and scientists who preferred more classical approaches always tried to emphasize. Since the previous jPCA approach only explained about 1/10 of the variance in the data, it was clear from the beginning that it cannot be the final truth and still it was used to claim that motor cortex would undergo simple rotational dynamics during a fast, transient reach, supporting the validity of the dynamical systems view. The new study tries to "rescue" the core of the idea. While it admits the much higher variability induced in the neural states between different reach conditions (acknowledging the fact that there might indeed be some form of "representational" information about the different reaches), by identifying different manifolds for each condition, it still emphasizes the rotational dynamcis within these manifolds.

The reviewer's summary of our approach is broadly correct, and highlights one of the central points of our paper. Both the representational and dynamical systems perspectives have much to offer, and here we are able to express the relationship between them. This both highlights the value of the representation, and helps us understand how the command signals are actually generated.

The former observation (condidition-dependent initial conditions) has been suggested and shown before (Perich et al. 2020; Michaels et al. 2020; acknowledged in the manuscript).

It is important to point out that our approach is not synonymous with condition-dependent initial conditions, though it does entail them. The original jPCA paper used condition-dependent initial conditions, describing them as the "seed" for the dynamical system so that the same system could produce different activity on different conditions. Here, we find that not only are the initial conditions condition-dependent, but the rotational planes themselves are condition-dependent too. We have now emphasized this distinction in the paper:

Introduction:

> The dynamical systems approach has further organized our understanding of population activity during reaching by showing that preparatory activity sets the initial state for future movement-epoch dynamics [19]…

Discussion:

> Based on our findings, we propose the following conceptual model of motor cortex during reaching (Fig. 8a). Unique to the LDR model, the population state of motor cortex is moved to the appropriate state space location, and the population state then rotates in the condition-specific planes set by the local dynamics around that state space location. As in many previous models, the phase and magnitude of the rotation are set by the initial activity, whether this is from a preparatory period or the irreducible preparation preceding movement [19,35,37].

So the important questions is: is the latter (rotational dynamics within the condition-specific planes) a non-trivial finding, what would be the alternative hypothesis, and which functional aspect of motor cortex do we understand better by this form of description?

We completely agree that this is a key question, and apologize that our answer was not clear enough in the original manuscript.

Regarding non-triviality, we might ask whether our factorization could describe any similarly smooth data at a similar level. We previously included the “recoverability” analysis (Fig. 3h) to show that our factorization described the data well and that the structure of the data meets several necessary conditions, but did not directly compare to ‘chance’. To address this, we now compare the variance captured of smoothed noise, including a p-value in Results and explanation in Methods. We also now better describe what structure the recoverability results require:

> More specifically, we expect recoverability to fail when two (or more) temporal basis functions are assigned the same dimension in state space by the loading matrix. As each temporal basis functions dynamically supports each other, this overlap would preclude neural activity from acting as a dynamical system.

The above analyses rule out triviality due to simple smoothness. But to address non-triviality more deeply, just as the reviewer notes we must consider the alternatives. There are several alternate hypotheses to condition-dependent rotational planes with conserved frequencies, including fixed rotational planes (as in jPCA), and neither conserved planes nor conserved rotational frequencies (as in other models of neural activity). We have added direct comparisons between the ability of these models to explain motor cortex activity, to supplement the previously existing explicit exploration of these models we already include in the text.

Finally, what do we learn about motor cortex? As we discuss more thoroughly in a reply above, we argue that we have better learned the nature of the dynamical system, discovered a class of dynamics that allows for much richer outputs and more flexible control than Churchland 2012-style planar rotational dynamics, and linked representation to dynamics. In our opinion, these significantly change how we view the system that is motor cortex, and allow us to understand it more deeply.

We had not previously made the implications of our findings clear enough, and have revised the text in several places to call out what we have learned about motor cortex as described more thoroughly in a reply above.

Since we are looking at a behavior that starts with a resting arm and ends with a resting arm, while briefly producing a pattern of muscle activations for reach in between, one wonders what the alternative is to neural dynamics that more or less (i.e. except for pose-dependend modulations) return to their initial conditions after running along a brief trajectory in neural state space? At the individual neuron level in motor cortex it is a well-known fact that many neurons start with no/low level of activity prior to reach and end with this level after the reach, showing transient, often biphasis activation during acceleration/decelaration in between.

We think there are really two questions here: Are dynamics in general trivial? And, given this specific behavior, could the appearance of dynamics be inherited from the behavior?

Regarding the non-triviality of dynamics, this has previously been litigated in the literature. Dynamics are not present in muscles whose activity is superficially similar to the neural responses in M1 (Churchland 2012), dynamics are stronger than expected given a highly-sophisticated shuffle that preserves not just the marginals but the interactions between marginals (Elsayed & Cunningham 2017), dynamics are much weaker in S1 (Russo 2018), and dynamics are surprisingly absent in hand M1 during a grasp task (Suresh 2020). The reviewer is correct that information should all be in the manuscript, and we have now added it:

> These reach-related dynamics are not trivial: they are stronger than expected from other aspects of neural activity [34], and are absent in muscle activity during reach [27], S1 during cycling [29], and hand M1 during grasp [35].

Regarding inheriting the structure of dynamics from the behavior, this question points out the need for a central control in the paper, which we did not previously describe in these terms. Here, we address it directly by showing that the preserved frequencies are not a consequence of the behavior. The relevant text in the Results now reads:

> The second concern is that a common time course could be present in motor cortex activity simply because different reaches take similar amounts of time, and the neural activity structure is inherited from the behavior. If similar reach time courses were the primary source of the similar neural frequencies across conditions, then warping reaches to identical durations should further improve this similarity. Prior to warping, the frequencies of the rotations were unrelated or weakly related to reach duration (M1-N, Pearson's Rho = -0.26, p = 0.023; other datasets, Pearson's Rho = -0.18-0.21, p > 0.067). Warping to equalize reach duration induced a negative correlation: the more a reach was warped, the less that condition's neural activity was fit by the common rotations (Fig. 3i; Pearson's Rho

> = -0.63 to -0.52, p < 0.001). The conserved rotational frequencies were therefore not explained by the similar time courses of the reaches.

One non-trivial finding probably is the observed conservation of eigen-frequencies in the rotational dynamics, since the authors say (page 3): "kinematic parameters performed poorly as predictors of eigenvalues". Unfortunately, it is not clear to me what the basis for this statement is (which analysis of the reach velocity profiles across different conditions shows this?) and what the functional interpretation/biological relevance of these preserved eigenvalues is?

Yes, this point was not highlighted sufficiently. We now reference the relevant analyses explicitly in the Results: "kinematic parameters performed poorly as predictors of eigenvalues (mean R^2 = 0.1, leave-one-out cross-validation; Extended Data Fig 2)." The interpretation is now covered briefly but directly in the Discussion:

> LDR may have other advantages for the brain, allowing motor cortex to generate richer command signals during movement than would be possible with strictly rotational dynamics. When limited to a single readout, planar rotational dynamics can approximate any arbitrary pattern over time given a well-chosen initial state. This allows rotational dynamics to drive the needed spiking in, for example, a single corticospinal neuron (CSN). Rotational dynamics, however, cannot arbitrarily set the phases and amplitudes of two or more CSNs' activity. Readouts from rotational dynamics are therefore strictly locked in phase and magnitude, meaning that spiking activity of multiple CSNs cannot be independently modulated without an explosion of model dimensionality (Fig. 8e). LDR does not have this limitation. With rotations oriented appropriately in state space, each CSN contains oscillations of the correct amplitude and phase to produce the needed pattern of spiking over time (Fig. 8f). By changing rotational planes between conditions, CSNs can be driven with effectively independent phases and magnitudes. Given that muscle activity for reaching can be assembled from a small basis set of sines and cosines [27,60], this makes LDR a potentially adequate generator for the required control signals. Note, however, that the data used here did not identify CSNs, and thus we cannot examine M1's outputs directly here.

The higher levels of explained variance per se, compared to other linear methods, is impressive, but I consider this an incremental improvement unless the here presented model explains something about motor cortex function that could not be unvealed with other methods, including non-linear methods. For example, in Figure 7, the encoding and decoding methods are not compared to other high-performing algorithms within the same domain modelling dynamical systems without the assumption of shared rotational frequencies, such as autoLFADS (Keshtkaran et al. 2022). Also, I think the method presented here would have to be demonstrated to generalize well to motor cortex data during different type of motor behavior.

We very much agree that the higher explained variance alone is not sufficient to warrant a high-profile venue. As described above, we do not consider this the central finding. Instead, we argue for the importance of these results based on the finding of a new class of dynamical system that better describes the neural data, the reconciliation with representation, the implications for how the brain solves the inverse problem, and the implications for how the brain can flexibly control and coordinate many degrees of freedom.

Regarding decoding, we think the previous version of the text was not clear about the point of these analyses. We did not intend to argue that our decoder would lead to superior performance in the brain-computer interface context vs. existing algorithms. Indeed, a hard lesson for the field in recent years is that such a claim requires using a decoder online, which we did not do and which this algorithm is not designed for. Instead, our motivation for using LDR-based decoding is that it demonstrates how strongly the location and orientation of rotations relate to the kinematic trajectory. This is a direct demonstration of the relevance of LDR to the inverse problem. Specifically, the brain could specify a simple static encoding of the reach trajectory in the inputs, this could push the location of PMd/M1 activity to the 'correct' location in neural state space, and the local dynamics would then generate the time-varying outputs. This scientific point is actually much clearer from the encoding models than the decoding models; to our knowledge, no previous encoding model has performed anywhere close to LDR-based encoding. This fact argues that we have gotten at the fundamentals of what these areas are doing in this task. However, all of these results are compatible with traditional decoding methods: in our model the outputs to the spinal cord are simply output-potent dimensions, as traditional decoders assume. With good recordings (many high-firing-rate neurons with plenty of tuning), traditional decoders should work well if our model is correct. LDR can take advantage of output-null activity, which can boost decoding when spike counts aren't huge, but there is no reason it should outperform traditional decoders in the high-SNR setting even if our model is exactly correct.

We have rewritten portions of the Results and Discussion to make this all clearer. Thank you for pointing out that we had led the reader in an unintended direction.

In Figure 8, the author asserts that their LDR method surpasses standard decoding by predicting activity in both the output-potent and output-null dimensions. However, none of the preceding figures differentiate between the potent subspace and null subspace, making the conceptual explanation disconnected from the remaining findings presented in the manuscript.

This is a great point. We have now explicitly implemented this analysis, and added Extended Data Figure 8, which demonstrates this point. For the reasons given above, we are also clearer that the superiority of this performance tells us about the brain but may or may not be helpful in a BCI context:

> Importantly, this decoding did not rely solely on neural activity in "output-potent dimensions" encoding for kinematics or muscle activity. Identifying and removing dimensions that encoded hand position, hand velocity, and muscle activity produced no substantial degradation in decoding quality (Extended Data Fig. 8; Wilcoxon Signed Rank test, p = 0.04-0.96)

Reviewer #2 (Remarks to the Author):

This is an interesting article that presents an ensemble of ambitious analyses of motor cortical activity during a large variety of reaches.
These extensive analyses shed new light on several aspects of motor cortical dynamics.
Specifically, the authors uncover that diverse reach conditions are associated with more diverse dynamical features - notably, oscillation planes and locations in neural state-space - than reported in previous reach studies. In addition, the authors discuss possible consequences for the expressivity and generalization abilities of motor cortical computations, which are of broad interest.

The article is already a valuable read, and I think that it can be further refined in a relatively straightforward manner by adding a few clarifications about the methodology used and its implications on how results are interpreted.

Thank you.

Main comments:

1. One main notion introduced in the manuscript is that of 'condition', as the authors compare various quantities across and within conditions. However, occasionally, authors also plot quantities colored by 'target angles' (for instance, for the eigenvalue analysis in extended data fig. 2).
After reading the manuscript, I still had some uncertainty about how 'conditions' and 'target angles' relate. The method says 'On some trials, the monkey was required to avoid virtual barriers presented at the same time as the target, eliciting curved reaches to produce 72 different reaching “conditions” (36 straight, 36 curved).'. Is one 'condition' several target angles but a single type of barrier and/or extent required? Or does a 'condition' refer to the combination of a single target angle, and a type of obstacle / reach extent required?

We apologize for the confusion, and have revised the manuscript to clearly define “condition”:

> Two monkeys, J and N, performed a “maze” variant of a delayed-reach task that evoked straight or curved reaches (Fig. 1a,b). We refer to the 72 unique combinations of target and virtual barrier positions as “conditions”.

As the reviewer notes, there are indeed analyses where we hold out groups of conditions based on target angle, simply to make cross-validation a bigger generalization challenge. We are now clearer about what we are doing in the relevant places.

2. The authors state that 'As previously shown, though, these rotational dynamics explained only 7-12% (s.d. < 13% across conditions; Extended Data Fig. 1) of the variance in peri-movement firing rates', and that this relates to a single LDS shared across conditions.

However, Lara, Cunningham and Churchland (Nature Communications, 2018) report that - across different reach angles - a single linear dynamical system fits the *reach-angle-dependent* part of M1 activity with an r-squared of R2 = 0.84 and 0.76 (and R2 = 0.74 and 0.62 when constraining the dynamical matrix to be skew-symmetric). I don't think that there is a direct comparison for this specific analysis in your case (you come close when using jPCA across conditions, but it is still a different analysis). Of course, your subsequent observation of condition-specific oscillation planes would suggest that a dynamical fit across conditions would be poor in your data. If that's true (and it'd be nice to directly test this, even if using a simpler analysis than in Lara et al. and just fitting a linear dynamical system across conditions), is it because you are considering a larger variety of conditions than in the Lara paper? Indeed, in the current paper you have obstacles, while Lara et al. didn't (the different reach angles are what they consider as 'conditions' during the movement period). Is the extent to which conditions differ in dynamics related to how much they differ in terms of kinematics and/or (inferred) EMGs? This could be a step towards providing evidence for the idea mentioned in the discussion that 'allowing the population state to rotate in substantially different planes for different reaches may allow motor cortex to produce a variety of different output signals across reaches'.

A clearer apples-to-apples comparison is a good idea. Lara’s R2’s, referred to by the reviewer, are the fits of the LDS to activity within the rotational planes identified by the LDS, not the variance explained in motor cortex activity. That is, these values only consider the activity in the jPCA planes, to see how dynamical the activity is in just those planes. And it is fairly dynamical in those planes. But those planes account for only a modest fraction of the overall variance: 7-12%, as we noted.

To make this all clearer, we have now added both of these quantities to facilitate comparison and to highlight the difference between these high fit numbers and the low variance explained. In addition, the reviewer was correct about the complexity of reaches impacting fit. Inspired by this comment, we have added analyses to demonstrate that the higher complexity of reaches in this dataset do worsen the fit of a single LDS across conditions (Extended Data Figure 1b).

3. The authors perform several advanced analyses to try to estimate quantities that can be related to models of motor cortical computations. This is a worthwhile but difficult undertaking, notably because the statistical methods used for estimation have unknown bias and variances, that could differ across the quantities of interest. To mitigate these issues, as a general guideline, it might be best to use statistical tools that are as directly related as possible to the question that one is asking. Notably, there are tools that can fit latent linear dynamical systems directly from PSTHs, an approach that also allows computing confidence intervals and cross-validation across different trials, without relying on a Poisson assumption - see for instance O'Shea, Duncker et al., bioRxiv 2022 (Method 9. Latent linear dynamical system model). In contrast, here, it appears that the authors first use a dimensionality reduction method (paragraph 'Fitting dynamics to a single condition' of the methods) prior to fitting the dynamics. I suspect that the two methods may give slightly different results. Specifically, given that with the former method dimensionality reduction is not done a priori but is rather a consequence of the rank of the fitted connectivity matrix, this method might be able to extract more precise estimates of eigenvalues and eigenvectors. It'd be

great to at least discuss how the method chosen to estimate the quantities of interest may affect the results.

We thank the reviewer for their insightful comment. Indeed, we were separating the step of dimensionality-reduction from fitting an LDS. Combining these steps could improve accuracy.

We carefully reviewed the approach pointed out by the reviewer. That particular approach assumes Gaussian noise (Method 9. Latent linear dynamical system model), and fits an LDS using expectation maximization. Unfortunately we do not know of an existing method that avoids the requirement to assume either Gaussian or Poisson noise. However, we took to heart the reviewer's point about combining steps to improve accuracy, and modestly improved on O'Shea and Duncker's method for this purpose. In particular, we now fit low-dimensional linear dynamical systems using reduced-rank regression, which makes the same assumptions as the highlighted method but has an analytic solution. As the reviewer hypothesized, this in fact marginally reduces the variance in the estimated eigenvalues between conditions, and is definitely more principled as well. We also now call out the noise assumption explicitly in the Methods: "We fit the LDS using reduced-rank regression to find the optimal low-rank LDS that explained maximum variance in that condition's neural activity, assuming Gaussian noise in firing rates."

4. Relatedly, it may be difficult to conclude whether eigenvalues are significantly different across conditions - indeed, the estimated magnitude of eigenvalue distances due to noise (fig. 3c, noise) is similar to the magnitude of the eigenvalue distances across the eigenmodes estimated from single-condition data (fig. 3a and h).
Comparing distances between pairs of eigenvalues introduces the difficulty of knowing which eigenvalue corresponds to which, and of relying on a Poisson noise assumption - which may contribute to the above observation.

We thank the reviewer for this thoughtful comment. First, we wish to clarify: our intended message here was that the differences in the eigenvalues found on different conditions was only barely larger than expected due to estimation noise. Clearly it would always be better to have lower estimation noise, but our conclusion is that the frequencies are highly conserved across conditions, almost to the limit of our ability to detect differences. This has now been clarified in the Results: "The rotational frequencies (specified by the eigenvalues) were nearly identical between conditions: the variation in eigenvalues between conditions was only slightly larger than the floor due to estimation noise (ROC-AUC = 0.51-0.58; Fig 3b,c). This slight variation in eigenvalues additionally contained little-to-no information about ongoing reaches: kinematic parameters performed poorly as predictors of eigenvalues (mean R2 = 0.1, leave-one-out cross-validation; Extended Data Fig 2). This demonstrates that rotational frequencies are approximately conserved between conditions."

Second, the reviewer makes an important point about our reliance on the Poisson assumption here. In light of this concern, we have implemented a second version of these controls, using separate estimates of the eigenvalues by partitioning trials in half, as suggested in the comments above and below. This control replicates the previous results, but without assuming Poisson statistics. We now report outcomes from both methods in the manuscript.

As an alternative to comparing distances between pairs of eigenvalues, it could be interesting to compare, between conditions, the estimated eigenvalue density quantifying the uncertainty in eigenvalue position due to trial sampling (as in O'Shea, Duncker et al., bioRxiv 2022). For instance, one could use a Kolmogorov–Smirnov test on the projection of eigenvalue densities along either the imaginary or real axis.
A similar approach could also be used for eigenvector angles, where the advantage would be to get rid of the Poisson assumption and instead directly estimate the uncertainty due to sampling.

The suggested analysis is clever, but we intended to ask a somewhat different question as described in our reply above. Here, any reasonable statistical test will report a difference between the random-sampling distribution and the empirical distribution because these distributions are, in fact, slightly different. Our intended point was that this difference is very small. To quantify how distinguishable these distributions are, we used an ROC-AUC. As per our expectations, these distributions are indeed distinguishable, but only barely so.

With regard to eigenvector uncertainty estimation, we do indeed use trial resampling for exactly the reason the reviewer describes (Fig. 4b-c).

5. Two parts of the discussion were unclear to me.
First, you mention that 'Smoothly varying the rotational plane may allow motor cortex to generate the “correct” activity patterns for new reaches by reaping the benefit of strong generalization due to local linearity'. On the one hand, in a (quasi)-linear regime, previous work (e.g. Lara et al., Nature Communications 2018; Vyas et al., Neuron 2018; Vyas et al., Neuron 2020; Logiaco et al., Cell Reports 2021) has emphasized smoothness and/or generalization properties of the dynamics when simply changing the initialization of the circuit to create movement variants. If this is what you are referring to, it would be nice to clarify. If it is not, it would be nice to clarify this alternative idea further because the cited references either do not refer to dynamical cases or to quasi-linear dynamics, so I do not understand what the authors mean.

Yes, our idea isn’t identical but is deeply related to those noted, and discussing it in context of those papers makes sense. The relevant section of the Discussion now reads:

> This low-dimensionality of conditions, combined with high neural dimensionality due to rotational plane variation, may allow motor cortex to generalize well while nevertheless being sufficiently expressive [51,52]. Smoothly varying the rotational plane may allow motor cortex to generate the “correct” activity patterns for new reaches by reaping the benefit of strong generalization due to local linearity [53,54], as previously argued in dynamical systems analysis [42,55,56] and models [33] of motor cortex.

Second, I am also confused about the discussion around corticospinal neurons (CSN). First, you could have several separate subnetworks with linear dynamics (possibly, each of them having rotational dynamics) that each project to a separate CSN or CSN population. This mechanism could separately modulate the phase and amplitude of each CSN, which seems to contradict your statement that 'Readouts from rotational dynamics are strictly locked in phase and magnitude, meaning that spiking activity of multiple CSNs cannot be independently modulated'. Second, given that your results concern a nonlinearity *across conditions, not across time* while within each condition the activity is well-described by linear dynamics, I do not understand how the nonlinearity could help control different CSNs that are simultaneously active during one condition. More generally, it could be good to discuss that one difficulty in linking the authors' observations to interpretations about M1's function is that the output of M1 is not directly experimentally accessible (we know it is a signal sent to the muscles, but this signal is only indirectly related to the recorded kinematics).

We agree that this topic warranted more careful discussion. The reviewer's first point is that dynamical systems can always be made more flexible by adding more dimensions. As we understand it, the reviewer's specific suggestion is equivalent to the system having dedicated dimensions for each CSN or group of CSNs. This is certainly true, but rather different from the previous models we are contrasting LDR against. In addition, dynamics in the CSN-specific modes would have to be minimally coupled to the dominant modes in the network. This is possible, but would probably mean that dynamical systems aren't a particularly useful way to describe how the system controls movement. As described in the previous reply (and following the excellent suggestions of this reviewer), we are now clearer in the Discussion that we are arguing for an intermediate point: more flexible than the simpler LDSs that are mainly discussed in prior literature, but still substantially constrained and therefore yielding the benefits of low-D noise robustness and generalization.

Regarding the second point, we illustrate the answer in Figure 8. Suppose that for a leftward reach you want a 2 Hz sine component in both the biceps and deltoid, and wish for them to start at the same phase. For a rightward reach, suppose you still wish to have a 2 Hz oscillation in both muscles, but want the phase 90˚ delayed in the deltoid relative to the biceps. With a standard LDS (or planar rotations in particular), this is not possible without mode doubling (repeated eigenvalues), which entails the problem described above - every muscle must now be controlled separately and you do not reap the benefits of a low-D system. In LDR, you simply tilt the 2 Hz plane so that different amounts of the sine and cosine components are in the deltoid's output projection, and thereby alter the phase. We have now tried to clarify this in the relevant part of the Discussion:

> With rotations oriented appropriately in state space, each CSN contains oscillations of the correct amplitude and phase to produce the needed pattern of spiking over time (Fig. 8f). By changing rotational planes between conditions, CSNs can be driven with effectively independent phases and magnitudes. Given that muscle activity for reaching can be assembled from a small basis set of sines and cosines [27,61], this makes LDR a potentially adequate generator for the required control signals.

Finally, the reviewer is right that we do not observe the activity of identified CSNs. It is therefore not possible from these data to know what output signals M1 produces, or needs to produce. However, LDR is strictly more flexible than planar rotational dynamics. This means that LDR enables more possibilities in what outputs can be produced, whatever is required. In the text above, we tried to navigate this distinction more carefully.

Additional remarks:

1. Is the fact that you are able to better decode kinematics from the LDR related to previous work that has shown how BMI decoding can benefit from accounting for the dynamical nature of M1 activity (e.g. Kao, Ryu and Shenoy, Scientific Reports 2017; Kao, Nuyujukian, Ryu, Churchland, Cunningham and Shenoy; Nature communications 2015)?

Yes, it is related though not the same point. We have changed the text to explain:

> Our findings suggest several immediate avenues of future research. LDR-based decoding, as a sequence-to-sequence model, cannot be used for real-time control of brain computer interfaces (BCI), except perhaps in tasks such as speech which may naturally be organized as sequence-to-sequence problems and where the relevant parts of motor cortex are known to exhibit dynamics [30]. Our findings argue for a new form of dynamics, which could be exploited for decoding in multiple ways. Previous methods have used dynamics to incorporate information from output-null dimensions to denoise the output-potent dimensions that are read out [62]; or, used whole neural trajectories to estimate the current one [60]. A better understanding of the dynamics may be able to improve performance with such methods. Alternatively, the location and orientation of the dynamics themselves might be used directly for decoding.

2. The discussion refers to 'autonomous' dynamics, but I am not sure why the argument matters here. If the authors mention it, it would be nice to clarify that there could really be many meanings behind this term. Notably, are the authors referring to the hypothesis that M1 is mostly driven by its own recurrent connections (as opposed to strongly interacting with other brain regions), or to the notion of feedforward control from the optimal control literature which differentiates whether or not the controller receives fast and reliable sensory feedback from the external world?

The reviewer is absolutely right. We removed this argument.

3. You are mentioning the danger of 'overfitting' when fitting dynamics. As mentioned above, there are ways to separate training and test sets to control for this possibility.

We thank the reviewer for highlighting the mentioned methods, which we have now used as an independent way to confirm our results. We have removed the mention of overfitting.

4. For Fig. 3h, you mention that 'The projections from each condition’s motor cortex activity recovered the temporal basis functions almost perfectly (91-95% variance explained)'. Does that refer to the mean projection across conditions? The variance around the individual translucent traces in Fig. 3h seems large, so I would expect the error computed per condition to be larger.

That statistics refer to each condition individually, not the variance between them. To quantify how much they vary across conditions, the translucent regions show standard deviations (not SEMs), which is why they appear large relative to what we are all used to seeing.

Reviewer #3 (Remarks to the Author):

Review of: Reach-dependent reorientation of rotational dynamics in motor cortex
The manuscript presented by Drs. Sabatini and Kaufman builds upon previous work describing rotational dynamics in primate motor cortical single unit ensembles. The critical innovation is expanding the model of rotational dynamics to examine multiple planes and frequencies. While frequencies remained relatively stable across conditions, the planes of rotation varied more than would be expected by chance. Modeling rotations in multiple planes (location dependent rotations, LDR) greatly increases the explanatory power of the resulting dynamical system models. Models with larger numbers of parameters are expected to explain greater degrees of variance (as the authors correctly point out in the results section). However, the increase in explained variance from < 12% to > 90% suggests that the new model is capturing important sources of variance ignored by previous formulations. Furthermore, standard models applied to individual conditions (single planes) also explain a high fraction of condition-specific variance. The results presented suggest that the increase in variance explained is tied specifically to adding additional rotation planes, but not frequencies (which are kept consistent in the final LDR model). I think this point could be made more explicit by making this comparison directly in a figure: I was expecting to find a figure comparing variance explained by taking into account rotations in a single plane (i.e. jPCA), an LDR model fitting the same frequencies in multiple planes (the presented model), and one fitting both planes and frequencies. Figure 2D presents part of this comparison by contrasting jPCA with single condition models, but it feels like it does not cover the full set of comparisons.

We thank the reviewer for their kind words, and agree that adding a graphic for this comparison is a great idea. We previously discussed the various alternatives in the text: fitting a model of each condition's dynamics individually with an LDS allows for both different rotational frequencies and planes. But the reviewer is right that this point deserved a figure. We have added an Extended Data Figure comparing variance accounted for by individual-condition LDS's (individual frequencies and planes), LDR (shared frequencies and individual planes), and jPCA (shared frequencies and planes). This shows clearly that the drop in variance explained by sharing planes is trivial, because the condition-specific models find the same frequencies in each condition anyway.

The authors introduce the Subspace Excursion Angles (SEA) metric, which orders subspaces according to the angles between them. Using this strategy they identify 6-20 distinct planes with angles > 45˚ to each other in motor cortex activity. The authors claim that "Existing methods of estimating dimensionality do not distinguish slight variations from more substantial variations, as long as the occupancy of each additional dimension is above the noise level." While I believe this is true of standard projection-based methods such as PCA, I am not certain it applies to fractal intrinsic dimensionality estimation methods. While the SEA method seems intuitive and practical, I think it might be useful to relate these results to more standard metrics. For references to possible alternative algorithms, I would suggest looking at:
1. Camastra, Francesco. "Data dimensionality estimation methods: a survey." Pattern recognition 36.12 (2003): 2945-2954.

2. Facco, E., d'Errico, M., Rodriguez, A. et al. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. Sci Rep 7, 12140 (2017).

We thank the reviewer for these ideas. The overall point is well taken: whenever you introduce a new metric it is important to cross-reference it with existing metrics to the extent possible. The particular metric cited by the reviewer, intrinsic fractal dimensionality, is a powerful method of describing the intrinsic dimensionality of a dataset, but in our case we are interested in the extrinsic dimensionality of the rotational planes, not the intrinsic dimensionality. To address the concern, we have supplemented SEA with several standard measures of extrinsic dimensionality and compared their results with SEA. In particular, we quantified the number of PCs required to capture 80% of each rotation's variance across conditions, the participation ratio of each rotation, and the number of dimensions occupied by each rotation with an SNR > 1 (see subsection titled **Rotational planes differed across reaches**).

The relationship between neural activity and movement kinematics is explored using a "sequence to sequence" (StS) encoder framework, i.e. relating single trial dynamics to the set of kinematics (summarized through dimensionality reduction) for a complete reaching motion. This approach can be used to make predictions in both directions. In the final section of the manuscript, the authors highlight the prediction of single trial kinematics from firing rates processed through LDR. The results presented in figure 7F suggest that LDR can greatly improve performance compared to standard instantaneous decoding methods. However, in the final paragraph of the discussion, the authors point out that StS encoders cannot be used for real-time control. This makes the comparison in 7F fall a bit flat. Comparing instantaneous prediction using limited time windows to predictions from one full sequence to another does not seem like a fair comparison, since less information is taken into account for each discrete estimate in the former case. It could be useful to add a comparison with some other kind of StS decoder (perhaps using a deep learning framework). I think this point should be addressed more fully in the discussion. Perhaps the authors could also speculate on ways incorporate StS decoding to BCI control. Maybe a complete reach to grasp action could be decoded (and even displayed in advance using augmented reality) before it is carried out by a robotic limb. Recent work in speech reconstruction could potentially be better suited for an StS approach, given the intrinsic sequential nature of language.

We apologize for being unclear about why the decoding methods were included, which has been pointed out by other reviewers as well. We did not mean to suggest the current LDR-based decoding as a "competitor" for online control. Rather, it was intended as a scientific proof-of-principle demonstrating the linear relationship between dynamics location and orientation to kinematics, which supports the point that this may be a core part of how the brain converts a static representation of movement plan to the time-varying control signals needed (i.e., helping solve the inverse problem). It also demonstrates that understanding a brain region's dynamics helps with decoding. We did not intend, however, to argue that this is likely to improve BCI control directly, which is a separate, large undertaking. Importantly, almost all the methods we compared against are similarly acausal, such as decoding from GPFA factors.

We have changed the text to make these points clearer, and to make clear that we are making a point about the relationship between motor cortex dynamics and kinematics, not LDR as a method of controlling BCIs. We have additionally expanded the relevant sections of the Discussion. Finally, we are also now clearer about what kinds of paths might lead to this science improving BCI, which now concludes our Discussion:

> Our findings suggest several immediate avenues of future research. LDR-based decoding, as a sequence-to-sequence model, cannot be used for real-time control of brain computer interfaces (BCI), except perhaps in tasks such as speech which may naturally be organized as sequence-to-sequence problems and where the relevant parts of motor cortex are known to exhibit dynamics [30]. Our findings argue for a new form of dynamics, which could be exploited for decoding in multiple ways. Previous methods have used dynamics to incorporate information from output-null dimensions to denoise the output-potent dimensions that are read out [62]; or, used whole neural trajectories to estimate the current one [60]. A better understanding of the dynamics may be able to improve performance with such methods. Alternatively, the location and orientation of the dynamics themselves might be used directly for decoding.

Overall, I think the presented work constitutes an impressive advance in our understanding of motor cortex encoding, presenting an elegant model that greatly expands the explanatory power afforded by previous work. If the three relatively minor points outlined above are addressed, I would fully support publication of the manuscript in Nature Communications.
Carlos Vargas-Irwin
Assistant Professor
Brown University Neuroscience Department
P.S. I also found what appears to be a Typo in the method section:
“We attempted therefore “recovered” the temporal basis functions from neural activity as
(6) $B'(c) = L(c)\dagger X(c)$
where $B'(c)$ is the recovered temporal basis functions for condition c and † indicates pseudo-inversion.”

The reviewer indeed found a typo. This has been corrected. We thank the reviewer again for his kind words and insightful comments.

REVIEWERS' COMMENTS

Reviewer #2 (Remarks to the Author):

I thank and congratulate the authors for performing new analyses that strengthen the paper, and for addressing several of my questions. The manuscript contains results of significance that should be communicated, and I think Nature Communications is an appropriate venue to communicate them. I also believe this updated version of the manuscript needs adjustments in some places to align the claims with the conclusions that can be drawn from the analyses. After this, I trust that the manuscript will be an influential contribution to the field.

1. Concerning point 2 in my initial review, I am very pleased to see that the authors have provided new analyses separating straight from curved reaches, the former being appropriate for comparison with some other works. However, it seems that there remains some confusion relative to the comparison with other analysis methods used in the literature. Previous papers went beyond the restrictions of the original jPCA fit from Churchland et al. 2012 - which was known to capture only a small portion of the full variance of the data. Notably, previous works fit M1 activity with a single linear dynamical system (LDS) across conditions, such as Lara et al. (Nat Comm 2018) or O'Shea and Duncker (Biorxiv 2022). Note that a single LDS is less restrictive than jPCA - or 'pure oscillations' - notably because the LDS' eigenvalues need not be purely imaginary.

Indeed, contrary to the assertions of the authors in their response ("[Lara et al.'s] values only consider the activity in the jPCA planes, to see how dynamical the activity is in just those planes"; and the authors' manuscripts lines 90-101), the HDR analysis from Lara and colleagues is very distinct from, and more general than, jPCA. Lara and colleagues specifically write "jPCA has two shortcomings given our present goals. First [...] we wish to make fewer assumptions regarding the form of dynamics. Second, the central motif predicted by motor-cortex network models includes both rotational dynamics and a condition-invariant shift of the neural state. [...] HDR optimizes jointly for all aspects of the hypothesized structure. In contrast jPCA employs PCA or dPCA and then seeks rotational structure, which could cause structure to be missed. Unlike jPCA, the present use of HDR does not focus on rotations per se, reducing concerns that the method imposes a particular form of dynamics. HDR is thus simultaneously more principled, more powerful, and more conservative that past approaches." To address the shortcomings of jPCA, this HDR analysis separates the activity into (i) a condition-independent signal (to try to capture the 'trigger-like' signal you described in Kaufman et al eNeuro 2016, and which is often modeled as an external input to a dynamical system - see e.g. in Zimnik and Churchland 2021); and (ii) a condition-dependent signal - which they show is very well-fit across different straight reaches with a single LDS. One may or may not find the methodology used by Lara et al. to extract a condition-independent signal justified - but then, it would need to be discussed for its own sake, instead of being lumped in the shortcomings of the planes identified by jPCA.

Along the same lines of successfully using non-jPCA-based dynamical models to fit motor cortical activity, OShea, Duncker & colleagues (biorXiv 2022, Method 9) fit a single LDS across straight reaches with piecewise constant inputs, and got 75-95% cross-validated variance explained (their Fig. 3 e-f). As I noted in my initial review, the ways that the two papers above fit the data across straight reaches with a single LDS, and the methods used in the current manuscript under consideration, are all a little different, which complicates comparing their outcomes. Notably, the other papers' models can capture slightly richer inputs to the dynamics while, as I understand it, you just allow real eigenvalues that translate into additive offsets with exponential timecourses. However, given the results in your extended data Fig. 1b, I am inclined to believe that one of the likely causes for your difficulty in fitting your dataset with a single LDS is higher task complexity (curved reaches in addition to straight reaches) compared to the previously fit straight reaches' activity.

In summary, it is clear that recent articles show that M1 activity *during straight reaches* can be very well fit using a single linear dynamical system (with a single effective connectivity matrix) - which notably involves initial conditions and/or piecewise constant inputs that are condition-specific, presumably leading to different state-space locations across conditions. In this context, the *important and new* results demonstrated by the authors are that (i) the quality of the fit of such models degrades when considering a more realistic and varied ensemble of reaches; and (ii) an alternative

model with condition-dependent effective connectivity captures the data very well. This finding appears to match well the prediction of a recent model (Logiaco et al, 2021) which showed that changing the effective connectivity across different conditions could be an efficient computational solution to increase the expressivity of M1 dynamics (more efficient than adding completely new neural populations/dimensions for each new condition).

I am worried that, right now, the paper reads as if jPCA (that the authors appear to use interchangeably with the phrase 'rotational dynamics', even though I am not 100% sure which exact assumptions this is referring to) is the current 'baseline' analysis used to characterize M1 activity. Further, the manuscript appears to use jPCA's poor ability to fit M1 data as a rationale for the new proposed model. At the same time, the author's results then actually contrast fitting data across conditions with a single linear dynamical system, to fitting different dynamical systems (with a focus on different

eigenvectors, but see below) across conditions. I believe that the authors should clarify that the latter comparison is the focus of their paper, and acknowledge previous literature that fits simpler reaches' M1 activity using a single LDS with inputs. This would allow the author's principal results to shine.

2. I congratulate the authors on using a more principled approach to fit their linear dynamics, and I am happy to see the changes implemented in the reported statistics in the text as well as the methods. However, I am puzzled that I do not see any perceivable changes in the corresponding Fig 3 a-c and extended data Fig 2. Maybe the former figures were included in the revised manuscript by mistake?

3. Concerning point 4 in my initial review, I wholeheartedly agree with the authors that it is important to focus on the effect size of eigenvalue changes as opposed to simply looking at whether changes are statistically significant. However, I feel that there was some misunderstanding, as my concern precisely focuses on effect size.

Looking at your Fig 3a left, I can see that the distance between the *within-condition* eigenvalues - corresponding to what you present as different 'true' basis functions - is of order 0.1 (e.g. see the distance between the eigenvalues corresponding to the 1.5Hz vs. 2.5 Hz basis functions). This means that 0.1 has to be considered a large and relevant effect size for eigenvalue distance: it corresponds to the difference between the blue and purple traces you put on display as separate basis functions in Fig 3h. It turns out that this large distance of 0.1 is also the average magnitude of eigenvalue distances not only across conditions, but also across the different partitions of within-condition trials used to quantify the estimation noise (Fig 3c). In some sense, a distance of 0.1 is taken to be a relevant ‘signal’ in Figs 3a and 3h, while it is shown to be the size of the noise fluctuations in Fig. 3c. In other words, eigenvalue estimates vary largely when considering different noisy realizations, as much as the relevant difference between the blue and purple traces of Fig. 3h. To me, this strongly suggests that eigenvalues are too difficult to estimate to conclude whether they meaningfully change or not across conditions.

Similarly, your new Extended Data Fig. 3 also strongly suggests that eigenvalues are very difficult to estimate in your data. Indeed, it shows that fitting LDS with condition-specific eigenvalues leads to worse r-squared than sharing the eigenvalues across conditions - even though the former has more free parameters.

This difficulty in estimating eigenvalues in your data resonates with recent theoretical results (Landau et al., PRE 2023) showing that the singular values of a matrix X (equivalently, eigenvalues of the matrix X X*) can be harder to estimate than its singular vectors (equivalently, eigenvectors of the matrix X X*).

Given these considerations - and given that, despite the large uncertainty in eigenvalue estimation discussed above, you do find some statistically significant differences across conditions as well correlations with reach parameters (extended data figure 2) - I do not currently see strong evidence that eigenvalues are conserved across conditions. I also started wondering whether choosing different frequencies for your fixed rotations - say 0.4 Hz, 2 Hz, 3.5 Hz and 4.25 Hz - could lead to a similarly great fit of your data as it would form a good enough general basis set. Along the same lines, given that your fitting procedure explicitly searches for directions in neural space that recover basis functions with your chosen frequencies, I wonder whether your finding that the correct frequencies are recovered for all trials (fig. 3h) could also be replicated for different chosen frequencies than those reported in the paper - as you would find a new loading matrix adjusted to yield these new frequencies.

While I have some reservations about the current manuscript’s conclusions concerning eigenvalues, I do not believe that this point diminishes the relevance of the authors' results. Indeed, the authors clearly demonstrate that some aspects of the dynamics vary largely across conditions when considering a variety of realistic and complex reaches, which I believe to be significant for the field of motor neuroscience. In addition, I believe that, by highlighting the difficulty of estimating eigenvalues from neural data, the authors' manuscript is also of technical value to the community.

Reviewer #3 (Remarks to the Author):

After reviewing the new version of the manuscript, I find that all of my suggestions have been incorporated. I believe the concerns raised by other reviewers have have also been adequately addressed. I fully support the updated version of the manuscript for publication in Nature Communications.

We thank reviewer 2 for their time, careful attention, and helpful comments. The reviewer requested that two remaining minor points of language be cleared up. First, they asked that the relationship of our work with several other papers on motor cortex dynamics be made clearer. Second, they raised a point about how we framed our ability to estimate variability in the eigenvalues of motor cortex dynamics. We have now addressed both these concerns, as described below.

Reviewer comments appear in blue, our replies appear in black, and quotations from the manuscript appear in red.

Reviewer #2 (Remarks to the Author):

I thank and congratulate the authors for performing new analyses that strengthen the paper, and for addressing several of my questions. The manuscript contains results of significance that should be communicated, and I think Nature Communications is an appropriate venue to communicate them. I also believe this updated version of the manuscript needs adjustments in some places to align the claims with the conclusions that can be drawn from the analyses. After this, I trust that the manuscript will be an influential contribution to the field.

We thank the reviewer for their kind words.

1. Concerning point 2 in my initial review, I am very pleased to see that the authors have provided new analyses separating straight from curved reaches, the former being appropriate for comparison with some other works. However, it seems that there remains some confusion relative to the comparison with other analysis methods used in the literature. Previous papers went beyond the restrictions of the original jPCA fit from Churchland et al. 2012 - which was known to capture only a small portion of the full variance of the data. Notably, previous works fit M1 activity with a single linear dynamical system (LDS) across conditions, such as Lara et al. (Nat Comm 2018) or O'Shea and Duncker (Biorxiv 2022). Note that a single LDS is less restrictive than jPCA - or 'pure oscillations' - notably because the LDS' eigenvalues need not be purely imaginary.

Indeed, contrary to the assertions of the authors in their response ("[Lara et al.'s] values only consider the activity in the jPCA planes, to see how dynamical the activity is in just those planes"; and the authors' manuscripts lines 90-101), the HDR analysis from Lara and colleagues is very distinct from, and more general than, jPCA. Lara and colleagues specifically write "jPCA has two shortcomings given our present goals. First [...] we wish to make fewer assumptions regarding the form of dynamics. Second, the central motif predicted by motor-cortex network models includes both rotational dynamics and a condition-invariant shift of the neural state. [...] HDR optimizes jointly for all aspects of the hypothesized structure. In contrast jPCA employs PCA or dPCA and then seeks rotational structure, which could cause structure to be missed. Unlike jPCA, the present use of HDR does not focus on rotations per se, reducing concerns that the method imposes a particular form of dynamics. HDR is thus simultaneously more principled, more powerful, and more conservative that past approaches." To address the shortcomings of jPCA, this HDR analysis separates the activity into (i) a condition-independent signal (to try to capture the 'trigger-like' signal you described in Kaufman et al eNeuro 2016, and which is often modeled as an external input to a dynamical system - see e.g. in Zimnik and Churchland 2021); and (ii) a condition-dependent signal - which they show is very well-fit across different straight reaches with a single LDS. One may or may not find the methodology used by Lara et al. to extract a condition-independent signal justified - but then, it would need to be discussed for its own sake, instead of being lumped in the shortcomings of the planes identified by jPCA.

We apologize for the confusion. We had added this comparison in response to a previous reviewer concern and did not include enough context. This comparison was intended to be against a secondary analysis in Lara 2018, not Lara's main analysis.

To be clear, there are two separate issues here.

First is that Lara 2018's main analysis using HDR is a pure subspace partitioning, not a dynamical system fit. That analysis isn't really relevant to our present work (because it isn't a dynamics fit), and we did not intend to invoke it.

Second, as the reviewer correctly points out, a single LDS is less restrictive than the low-dimensional rotational LDS fit by jPCA, and other work has used LDSs to model motor cortex dynamics. To make this latter comparison directly, and to avoid any further confusion, we have supplemented our analyses by fitting an unconstrained, full-dimensional LDS directly to peri-movement motor cortex activity. This procedure can account for only approximately half the variance in motor cortex activity, which we now report in our manuscript. Exactly as the reviewer notes (and which we analyzed at their excellent suggestion last round), this is at least partly because our dataset includes more complex reaches than those datasets.

We hope this additional analysis makes clear that our work is not motivated by a specific failing of jPCA, but the broader limitations of LDS models in explaining motor cortex dynamics. We have revised the relevant text, which now reads:

> This limitation in variance explained was partly due to using too-low dimensionality and an overly-constrained dynamical system. When including very small numbers of straight reaches, a single linear dynamical system (LDS) can indeed fit most of motor cortical activity [43]. The data used here, however, included a much wider variety of straight and curved reaches. A single LDS on these data only captured an average of 46-66% of the population variance (s.d. < 16%). This argues that linear dynamics, and rotational dynamics in particular, are incomplete models of activity in motor cortex.

Along the same lines of successfully using non-jPCA-based dynamical models to fit motor cortical activity, OShea, Duncker & colleagues (biorXiv 2022, Method 9) fit a single LDS across straight reaches with piecewise constant inputs, and got 75-95% cross-validated variance explained (their Fig. 3 e-f). As I noted in my initial review, the ways that the two papers above fit the data across straight reaches with a single LDS, and the methods used in the current manuscript under consideration, are all a little different, which complicates comparing their outcomes. Notably, the other papers' models can capture slightly richer inputs to the dynamics while, as I understand it, you just allow real eigenvalues that translate into additive offsets with exponential timecourses. However, given the results in your extended data Fig. 1b, I am inclined to believe that one of the likely causes for your difficulty in fitting your dataset with a single LDS is higher task complexity (curved reaches in addition to straight reaches) compared to the previously fit straight reaches' activity.

We agree. Crucially, in the mentioned study, monkeys only performed 4 straight reaches. This low condition count is almost certainly sufficient to explain the high variance explained by an LDS model. To confirm this, we generated 100 datasets of 4 randomly-chosen straight reaches, and fit LDS to the corresponding neural activity. In 86% of these datasets, the LDS explained >80% of the population variance. Again, we thank the reviewer for suggesting that we analyze straight and curved reaches separately to confirm the importance of more complex movements.

In summary, it is clear that recent articles show that M1 activity *during straight reaches* can be very well fit using a single linear dynamical system (with a single effective connectivity matrix) - which notably involves initial conditions and/or piecewise constant inputs that are condition-specific, presumably leading to different state-space locations across conditions. In this context, the *important and new* results demonstrated by the authors are that (i) the quality of the fit of such models degrades when considering a more realistic and varied ensemble of reaches; and (ii) an alternative model with condition-dependent effective connectivity captures the data very well. This finding appears to match well the prediction of a recent model (Logiaco et al, 2021) which showed that changing the effective connectivity across different conditions could be an efficient computational solution to increase the expressivity of M1 dynamics (more efficient than adding completely new neural populations/dimensions for each new condition).

Thank you.

I am worried that, right now, the paper reads as if jPCA (that the authors appear to use interchangeably with the phrase 'rotational dynamics', even though I am not 100% sure which exact assumptions this is referring to) is the current 'baseline' analysis used to characterize M1 activity. Further, the manuscript appears to use jPCA's poor ability to fit M1 data as a rationale for the new proposed model. At the same time, the author's results then actually contrast fitting data across conditions with a single linear dynamical system, to fitting different dynamical systems (with a focus on different eigenvectors, but see below) across conditions. I believe that the authors should clarify that the latter comparison is the focus of their paper, and acknowledge previous literature that fits simpler reaches' M1 activity using a single LDS with inputs. This would allow the author's principal results to shine.

We thank the reviewer for their astute observations. As we discuss above, after careful consideration we believe our new analysis and edits to the manuscript address the reviewer’s concerns.

2. I congratulate the authors on using a more principled approach to fit their linear dynamics, and I am happy to see the changes implemented in the reported statistics in the text as well as the methods. However, I am puzzled that I do not see any perceivable changes in the corresponding Fig 3 a-c and extended data Fig 2. Maybe the former figures were included in the revised manuscript by mistake?

We had in fact updated the figures with the new method, but as the reviewer correctly notes, the figures have not perceivably changed. In practice, the previous method for fitting single-condition LDSs produced almost identical results as the more principled method we adopted after the reviewer's previous comments. Both methods work by fitting an LDS and projecting it into the top few principal components of the data. Previously, we were estimating these quantities separately, whereas reduced-rank regression (our current approach) estimates them simultaneously. These quantities, however, can be estimated quite stably from the current data (even when done separately), meaning that numerically the two approaches yield almost identical results here. Nevertheless, we have opted to continue using reduced-rank regression as previously suggested, as it is theoretically a more justified method and we hope that other researchers pursuing our approach will follow.

3. Concerning point 4 in my initial review, I wholeheartedly agree with the authors that it is important to focus on the effect size of eigenvalue changes as opposed to simply looking at whether changes are statistically significant. However, I feel that there was some misunderstanding, as my concern precisely focuses on effect size.
Looking at your Fig 3a left, I can see that the distance between the *within-condition* eigenvalues - corresponding to what you present as different 'true' basis functions - is of order 0.1 (e.g. see the distance between the eigenvalues corresponding to the 1.5Hz vs. 2.5 Hz basis functions). This means that 0.1 has to be considered a large and relevant effect size for eigenvalue distance: it corresponds to the difference between the blue and purple traces you put on display as separate basis functions in Fig 3h. It turns out that this large distance of 0.1 is also the average magnitude of eigenvalue distances not only across conditions, but also across the different partitions of within-condition trials used to quantify the estimation noise (Fig 3c). In some sense, a distance of 0.1 is taken to be a relevant 'signal' in Figs 3a and 3h, while it is shown to be the size of the noise fluctuations in Fig. 3c. In other words, eigenvalue estimates vary largely when considering different noisy realizations, as much as the relevant difference between the blue and purple traces of Fig. 3h. To me, this strongly suggests that eigenvalues are too difficult to estimate to conclude whether they meaningfully change or not across conditions.

Again, we apologize for confusion here. The way we previously reported the within-condition spread was *summed* over the distances between all 7 or 9 corresponding eigenvalues. The relevant comparison with the distance between eigenvalues would be the *average* of the distances, not the sum. To avoid this confusion for future readers, we now report the average, which is (of course) 7-9 times smaller than what we previously reported. This means that the within-condition is an order of magnitude smaller than the differences between distinct eigenvalues. We have also updated Figure 3 accordingly. This all now makes it clearer that the variability due to estimation noise is quite small relative to the differences between eigenvalues. Thank you for identifying this issue.

Similarly, your new Extended Data Fig. 3 also strongly suggests that eigenvalues are very difficult to estimate in your data. Indeed, it shows that fitting LDS with condition-specific eigenvalues leads to worse r-squared than sharing the eigenvalues across conditions - even though the former has more free parameters.

We understand where this conclusion comes from, but the reason for this surprising difference is actually somewhat subtle. As the reviewer suggests, part of this difference is indeed likely due to numerical stability. But, as we quantify above, the influence of instability is actually rather small. If the eigenvalues had non-trivial differences, the small improvement in eigenvalue estimation due to LDR's pooling would be outweighed by using the wrong eigenvalues. This result therefore argues for even *better* conservation in eigenvalues than we can estimate. In addition, there is a second source of the observed difference: LDS models optimize fits of the state's derivative, not the state itself. LDR does optimize variance of the state itself.

This is all now better explained in the legend of what is now Supplementary Figure 3:

> Note that while the “different planes, shared eigenvalues” (LDR) model is a subset of the “different planes and eigenvalues” (condition-specific LDS) model, the “different planes, shared eigenvalues” explains greater neural variance when cross-validated. This improvement has two sources. First, as this model assumes that eigenvalues are shared across conditions, it gets to estimate rotational frequencies using every condition, leading to more stable estimates of eigenvalues. Second, LDS models predict the population state's derivative from the state, which does not directly optimize variance explained. The “different planes, shared eigenvalues” model, on the other hand, directly optimizes the variance explained.

This difficulty in estimating eigenvalues in your data resonates with recent theoretical results (Landau et al., PRE 2023) showing that the singular values of a matrix X (equivalently, eigenvalues of the matrix X X*) can be harder to estimate than its singular vectors (equivalently, eigenvectors of the matrix X X*).

Given these considerations - and given that, despite the large uncertainty in eigenvalue estimation discussed above, you do find some statistically significant differences across conditions as well correlations with reach parameters (extended data figure 2) - I do not currently see strong evidence that eigenvalues are conserved across conditions.

We believe the above clarifications address this issue: there was a miscommunication about the difference in scale between intra- and inter-eigenvalue distances, and the model cross-validation similarly argues that pooling eigenvalues is a very good approximation.

I also started wondering whether choosing different frequencies for your fixed rotations - say 0.4 Hz, 2 Hz, 3.5 Hz and 4.25 Hz - could lead to a similarly great fit of your data as it would form a good enough general basis set. Along the same lines, given that your fitting procedure explicitly searches for directions in neural space that recover basis functions with your chosen frequencies, I wonder whether your finding that the correct frequencies are recovered for all trials (fig. 3h) could also be replicated for different chosen frequencies than those reported in the paper - as you would find a new loading matrix adjusted to yield these new frequencies.

We think there are two possible concerns here. First is the more serious concern that the temporal structure of the data (the data's autocorrelation, for example) causes the data to be bandwidth-limited, meaning any well-chosen set of temporal basis functions within a certain frequency range would describe the data equally well. We have in fact controlled strongly for this possibility but did not emphasize this important point. We have amended our manuscript to do so. The relevant section now reads:

> In agreement with our fits to single conditions, the optimal temporal basis functions contained 3-4 rotations at 0.5, 1.5, 2.5, and 4 Hz, along with an offset (Fig. 3g; optimal number determined by cross-validation). These rotations explained 90-96% of the population variance (s.d. < 2% across conditions). To address concerns that this high-variance explained was due to smoothing, pre-processing, limited temporal bandwidth, or frequency-limiting artifacts, we shuffled time bins to disrupt temporal structure in motor cortex activity while preserving inter-unit spike correlations, before identically smoothing and pre-processing the shuffled data (Methods). This shuffle significantly lowered the variance explained by this method (35-46%; Wilcoxon Rank Sum Test, p < 0.001).

Second is the possibility that we could have chosen other frequencies for our rotations and gotten similar results. The answer to this one is more nuanced. Most importantly, we first point out that the core step in our matrix decomposition is just an SVD. There are no choices involved, and SVD is provably optimal in accounting for variance. Our main kinematic decoding operates on the SVD basis, without any further complications. For decoding, then, the question is moot.

When we do wish to find discrete frequencies, we fit an LDS to the basis functions, eigendecompose the LDS, then perform the non-orthogonal projection of the basis that the eigenvectors specify. All of these steps simply "clean up" the basis functions via projection to try to segregate different frequencies. This will do the provably-optimal job of extracting the frequencies that can be "purified." We further, in the Supplement, prove that this procedure finds the correct eigenvalues when the data is in fact generated by conserved eigenvalues, with eigenvectors varying between conditions. We note, though, that for numerical reasons they still do not end up perfectly pure, because the ultimate operation is simply a projection. Whatever other frequencies are present before the projection will remain after the projection. What this whole process does, then, is a compromise: it tries to purify frequencies to the extent possible without having the projection be so far from orthogonal that noise is unduly amplified.

So, could we choose different frequencies to recover? A different projection that did its best to purify a different set of frequencies would presumably work, but for the reasons above cannot work as well as what we did. Either the frequencies will be less pure, or the projection basis will be closer to singular.

Critically, though, this is not to say that the frequencies we present are an arbitrary compromise. This is suggested by Supplementary Figure 3, following the revisions we made at this reviewer's suggestion. Choosing to pool the eigenvalues fits the data better than optimizing them for each condition separately. This argues that there are "real" frequencies to be found, and therefore pooling to produce better estimates helps.

While I have some reservations about the current manuscript's conclusions concerning eigenvalues, I do not believe that this point diminishes the relevance of the authors' results. Indeed, the authors clearly demonstrate that some aspects of the dynamics vary largely across conditions when considering a variety of realistic and complex reaches, which I believe to be significant for the field of motor neuroscience. In addition, I believe that, by highlighting the difficulty of estimating eigenvalues from neural data, the authors' manuscript is also of technical value to the community.

We thank the reviewer for their thoughtful observations and time, and agree that these points were likely to come up for many readers. We think our amendments to the manuscript both address the reviewer's concerns and strengthen the manuscript.