A MULTIVARIATE GAUSSIAN PROCESS FACTOR MODEL FOR HAND SHAPE DURING REACH-TO-GRASP MOVEMENTS

Author(s): Lucia Castellanos, Vincent Q. Vu, Sagi Perel, Andrew B. Schwartz and Robert E. Kass

Source: *Statistica Sinica*, Vol. 25, No. 1 (January 2015), pp. 5-24
Published by: Institute of Statistical Science, Academia Sinica
Stable URL: https://www.jstor.org/stable/24311001
Accessed: 12-11-2020 14:27 UTC

Statistica Sinica **25** (2015), 5-24
doi:http://dx.doi.org/10.5705/ss.2013.226w

A MULTIVARIATE GAUSSIAN PROCESS FACTOR MODEL FOR HAND SHAPE DURING REACH-TO-GRASP MOVEMENTS

Lucia Castellanos[1], Vincent Q. Vu[2], Sagi Perel[1], Andrew B. Schwartz[3] and Robert E. Kass[1]

[1]*Carnegie Mellon University,* [2]*Ohio State University and* [3]*University of Pittsburgh*

Abstract: We propose a Multivariate Gaussian Process Factor Model to estimate low dimensional spatio-temporal patterns of finger motion in repeated reach-to-grasp movements. Our model decomposes and reduces the dimensionality of variation of the multivariate functional data. We first account for time variability through multivariate functional registration, then decompose finger motion into a term that is shared among replications and a term that encodes the variation per replication. We discuss variants of our model, estimation algorithms, and we evaluate its performance in simulations and in data collected from a non-human primate executing a reach-to-grasp task. We show that by taking advantage of the repeated trial structure of the experiments, our model yields an intuitive way to interpret the time and replication variation in our kinematic dataset.

Key words and phrases: Dynamical factor analysis, experiment structure, multivariate Gaussian process, reach-to-grasp, registration, variance decomposition.

1. Introduction

Accurate description of the variability in finger movement is central to understanding nervous system production of manual dexterity. In addition, characterization of finger motion variability is critical to the successful engineering of brain-computer interaction devices, where the goal is to provide individuals who have lost a limb with the ability to control a prosthetic hand. The human hand is enormously flexible but also hard to model because it contains over 20 degrees of freedom, mechanical constraints and, plausibly, complex and non-linear interactions among its components. There is much variability among subjects and even when the same subject performs the same grasp, two replications present different multidimensional trajectories. Part of the variability may be understood as a result of the constraints among the fingers, and this has led to the use of lower-dimensional representations known as *synergies* Santello, Flanders, and Soechting (1998); Todorov and Ghahramani (2004); Mason, Gomez, and Ebner (2001); Mason et al. (2004); Soechting and Flanders (1997); Pesyna, Pundi, and Flanders (2011); Thakur, Bastian, and Hsiao (2008). Standard matrix factor-

ization approaches, like principal components analysis (PCA), can go far in this direction but do not conform to the repeated-trial structure of most experiments and, furthermore, confound temporal variability with experimental condition and kinematic variability. In this paper we adapt a Multivariate Gaussian Process (MGP) model, also known as the Gaussian Process Factor Analysis model Yu et al. (2009), to decompose finger motion into two terms: a term that is shared among all replications of the same reach-and-grasp task and a term that is particular to each replication and that is modelled with a MGP. This provides a dynamic lower-dimensional representation of finger motion.

A *Macaca mulatta* monkey (considered a model for human hands) was trained to reach and grasp eight different objects presented in different orientations and spatial locations (Figure 1 of the Supplementary Material). The monkey was comfortably seated in a primate chair, with one hand restrained and the other free to move to perform the task. A state-of-the-art motion tracking system (Vicon Inc) was used to record the three-dimensional (3-D) positions of passive markers placed on a thin custom made glove worn by the monkey, at a rate of 200Hz. The markers were positioned at the center of each of the fingers' phalanges, on the wrist, and on the back of the hand. Each replication of the reach-to-grasp task corresponded to a specific condition (i.e. an object presented in a specific orientation) and constituted a multivariate time series of markers' position. The replicated reaches evolved across time somewhat differently on each trial, which poses a challenge: because important features of the data occur at different times on different trials, they could get lost when examining trial-averaged effects. In this paper we align (register) trials before applying our model, and we study the benefit of doing so.

The two main methodological contributions of our work include the alignment (or registering) of the collected multivariate grasping data and the decomposition and reduction of the dimensionality of the variation of the multivariate functional data according to the experimental structure: time, replication, and condition through the fitting of our Multivariate Gaussian Process Factor Model (MGPFM).

There have been other approaches in the literature to obtain temporal grasping synergies. For instance, Vinjamuri et al. (2007, 2010a,b) inspired by (d'Avella and Bizzi, 2005), proposed two convolved-mixture models that use SVD and an optimization step in their core, to learn a dictionary of time-varying synergies. While the approach is able to describe time varying phenomena, it does not provide a generative model of grasping. State-space models are generative models that have been used to model dynamics of general body motion Wang, Fleet, and Hertzmann (2008) albeit not finger dynamics. In these models, a Markovian assumption is posited and thus longer range time correlations are unable

to be directly captured. The model we present in this paper is not a state-space model—instead, we assume that the observed trajectories are generated via low-dimensional latent factor trajectories that are drawn directly from a GP, allowing for longer range correlations to be captured. Fyshe et al. (2012) also used this idea of modeling latent factor trajectories with GPs to analyze brain imaging data (MEG). As in Fox and Dunson (2011), Fyshe et al. (2012) assumed a loading matrix that changes over time as well as latent factors, which are correlated through a hierarchical Bayesian prior, while our model follows the more traditional setting of latent factor analysis by assuming that the factors are independent, and a stationary loading matrix.

In the following sections we introduce notation and formally explain our Multivariate Gaussian Process Factor Model. In addition, with views to clarity and as a fast reference, we included three summarizing tables in the Supplementary Material: Table 1 lists and defines the symbols used throughout the paper including observed variables, latent variables, and parameters being inferred; Table 2 enumerates the assumptions made on the parameters, and Algorithm 1 summarizes and describes our whole approach.

2. Model

Consider the p-dimensional observed dataset $\{Y_i^r(t) \mid i = 1,\dots,p;\ t = 1,\dots,T;\ r = 1,\dots,R\}$, where $Y_i^r(t)$ is the i^{th} coordinate at time t of the p-dimensional trajectory that is the r^{th} replication of an event. For simplicity, we consider only finitely many time points, but our model is based on Gaussian Processes and thus it applies to the continuous setting. Here R is the number of repeated trials, T the number of time slices, and p the number of observed variables. In our application the observed variables describe the hand kinematics – they could be position, velocity, acceleration, joint angles or any function or representation of hand kinematics.

The Multivariate Gaussian Process Factor Model (MGPFM) assumes

$$\begin{bmatrix} Y_1^r(t) \\ \vdots \\ Y_p^r(t) \end{bmatrix} = \begin{bmatrix} \mu_1(t) \\ \vdots \\ \mu_p(t) \end{bmatrix} + \begin{bmatrix} \sum_{j=1}^d b_{1j} X_j^r(t) \\ \vdots \\ \sum_{j=1}^d b_{pj} X_j^r(t) \end{bmatrix} + \begin{bmatrix} \epsilon_1^r(t) \\ \vdots \\ \epsilon_p^r(t) \end{bmatrix}, \tag{2.1}$$

where $\mu_i(t) \quad i = 1,\dots,p$ are deterministic mean functions, and $\mathbf{B} = (b_{ij}) \in \mathbb{R}^{p\times d}$ is a deterministic factor loadings matrix whose columns correspond to the d latent factors and rows correspond to the p observed variables. Each latent factor trajectory X_j^r is drawn iid from an MGP with mean function $\mathbf{0}$ and covariance function $\sum(t_1,t_2)$ defined by $\sum(\cdot;\cdot) : [0,1]\times[0,1] \to \mathbb{R}$, $\epsilon_i(t) \quad i = 1,\dots,p$ are

iid stationary MGP draws with covariance function $\Psi(t_1, t_2)$, which we assume to be diagonal in this work.

Letting $\mathbf{Y}^r(t) = [Y_1^r(t) \cdots Y_p^r(t)]^\intercal$, $\boldsymbol{\mu}(t) = [\mu_1(t) \cdots \mu_p(t)]^\intercal$, $\mathbf{X}^r(t) = [X_1^r(t) \cdots X_d^r(t)]^\intercal$, and $\boldsymbol{\epsilon}^r(t) = \left[\epsilon_1^r(t) \ldots \epsilon_p^r(t)\right]^\intercal$, (2.1) becomes:

$$\mathbf{Y}^r(t) = \boldsymbol{\mu}(t) + \mathbf{B}\mathbf{X}^r(t) + \boldsymbol{\epsilon}^r(t). \tag{2.2}$$

To ensure identifiability of the model we assume that $\mathbf{B}^\intercal\mathbf{B}$ is diagonal.

Our model (2.2) decomposes each kinematic trial into a term $\boldsymbol{\mu}(t)$ that is common among replications and a term $\mathbf{X}^r(t)$ that is specific to the replication. The spatial structure of the markers is encoded in the $\mathbf{B}$ matrix by summarizing and mapping down the spatial configuration of the hand to a lower dimension.

Parameter $\boldsymbol{\mu}(t)$ does not depend on the specific trial and can be modelled in two ways: invariant in time as a p-dimensional constant vector, and as a p-dimensional varying function in time. In the latter case $\boldsymbol{\mu}$ can be represented as a $p \times T$ matrix, or more efficiently, through a B-spline basis. The number of parameters to estimate for $\boldsymbol{\mu}$ is p when $\boldsymbol{\mu}$ is assumed to be constant, $p \cdot T$ when $\boldsymbol{\mu}$ is allowed to vary *freely* (with no constraints) as a function of time as $\boldsymbol{\mu} = \boldsymbol{\mu}(t) \in \mathbb{R}^{p \times T}$, and $\mathcal{O}(p \cdot c)$ when $\boldsymbol{\mu}$ is described through a $\mathcal{B}$-spline basis with c the number of basis functions where $c << T$. While the choice of basis is not scientifically important, our formulation drastically reduces the number of parameters to be estimated and it allows us to model the part of the variability that is common between trials in a smooth fashion.

The parameter Σ corresponds to the covariance matrix of the MGP. The form of Σ determines the properties of the low dimensional kinematic representation. Estimating $\Sigma \in \mathbb{R}^{T \times T}$ implies learning T^2 parameters. While it is possible to estimate with no constraints (which we do in some of our analyses), this procedure is prone to overfitting when there is not much data available or when the observed data was not drawn from the model. One way to overcome this problem is to impose structure to Σ by assuming that it takes a parametric form. In this work, we use an stationary exponential covariance function: $\Sigma(i, j) = \exp(-(i - j)^2/\theta_\Sigma)$ where θ_Σ controls the width of the diagonal that decays exponentially. But other functions such as the Matern covariance function are also possible Rasmussen and Williams (2006). In our case, the exponential covariance function effectively imposes a prior belief that the latent trajectories are smooth, where θ_Σ controls how fast the function varies within a certain window of time.

2.1. Estimation and inference problems

Our estimation algorithms are EM-based. We iterate between parameter estimation and inference on the latent variables. We refer to the Supplementary Material for derivation.

Since the learning algorithm is an EM procedure, it is susceptible to local optima. To avoid getting stuck in local optima we propose two initialization methods. The first (*MLE*) involves estimating the MLE for a matrix normal distribution Dawid (1981); Dutilleul (1999); and the second (*down-projection*) estimates the MGPFM parameters assuming a higher latent dimension and projects down to the desired dimension. For details see the Supplementary Material.

2.2. Identifiability

We assume orthogonality of the columns of $\mathbf{B}$, an identity covariance for $\mathbf{X}^r(t)$, and a covariance proportional to the identity matrix for $\boldsymbol{\epsilon}^r(t)$. This is enough to ensure identifiability of the parameters, because the MGPFM is completely specified by $\mathbb{E}[\mathbf{Y}^r(t)] = \mu(t)$ and $\mathbb{C}ov[\mathbf{Y}^r(t), \mathbf{Y}^r(u)] = \mathbf{B} \cdot \Sigma(t,u) \cdot \mathbf{B}^\mathbf{T} + \Psi(t,u)$. We assume (Table 2 in the Supplementary Material) $\Sigma(t,t) = 1$ and $\Psi(t,t) = \rho \cdot I_{p\times p}$ where $I_{p\times p}$, is the identity matrix of dimension $p \times p$. Then the equation of the variance is fully determined: $\mathbb{V}ar[\mathbf{Y}^r(t)] = \mathbb{C}ov[\mathbf{Y}^r(t), \mathbf{Y}^r(t)] = \mathbf{B} \cdot \mathbf{B}^\mathbf{T} + \rho \cdot I_{p\times p}$.

2.3. Variability of predictions and estimates

The main focus of this work is to provide a method for separating variation that is common across trials from variation that is specific to individual trials. We use the model as a tool to facilitate this decomposition; we use cross-validation to assess its adequacy in explaining the variation and predictive uncertainty, which has the benefit of not requiring the assumption that our model is correct. Alternatively, to assess uncertainty about the estimated parameters or its predictions without cross-validation, one can assume that the model parameters are known. In this case, the posterior distribution of the latent variables is Gaussian, and the uncertainty of the predictions can be assessed by inspecting the covariance matrix for the posterior distribution (Equation S2.13 of the Supplementary Material). However, the validity of this inference hinges critically on the correctness of the model and in this work we do not assume that our model is correct; instead we are using it as a tool to decompose the variability of the data.

3. Alignment Procedure

In functional data there are two types of variability: amplitude variation and phase variation Ramsay, Hooker, and Graves (2009); Ramsay and Silverman (2005). Amplitude variation describes the variability of the sizes of the features of the kinematic curves; features such as the height of peak velocities of different markers recording finger movement during a grasping task; phase variation describes the variability of the timings of such features of the kinematic curves

as the variation between the timing of the opening and closing, or between the peak velocities of the fingers.

We are interested in studying amplitude variations in order to understand how the movement of the fingers relate to one another. To isolate the amplitude variation from the phase variation we can transform the time-axis for each trial so that phase variation between kinematic curves is minimized.

Alignment of the R kinematic curves $\mathbf{Y}^r(t)$ is accomplished by estimating monotonically increasing *time warping functions* $h^r(t)$ such that the phase variation of $t \to \mathbf{Y}^r(h^r(t))$ is minimized across trials r. Note that it is important that the same function be used across kinematic variables for a fixed condition and trial, because we want to preserve the relationships between kinematic variables at fixed times. At the same time, different warping functions need to be learned for different replications.

Total energy signal. Alignment of multivariate curves is greatly simplified by summarizing each multivariate curve by a univariate curve. We summarize the trials based on the *total energy signal* based on the velocity of the markers. The basic idea is that the velocity of the markers typically has clear peaks, valleys, and zero crossings – features that are easily identified.

In order to describe the alignment procedure in a clear and compact way we slightly overload notation in this Section and in the Supplementary Material (S4.1) by denoting the original kinematic marker positions at time t as a matrix (instead of a vector). In particular, we rearrange the elements in $\mathbf{Y}^r(t)$ and consider the matrix $\mathbf{Y}^r(t) \in \mathbb{R}^{K\times 3}$ where $p = 3 * K$ and $K = 16$ is the number of markers placed on the fingers (K is multiplied by three because of the 3-dimensional positions of the markers). Likewise, $\dot{\mathbf{Y}}^r(t) \in \mathbb{R}^{K\times 3}$ refers to the corresponding velocities. Thus $\mathbf{G}^r(t) = [\dot{\mathbf{Y}}^r(t)][\dot{\mathbf{Y}}^r(t)]^\mathtt{T}$ is the matrix of inner products of marker velocities for each replication r in a specific condition, and we define the total energy signal of a trial as the sum of the squared magnitudes of the velocities across markers, which can be written as:

$$E^r(t) = \mathtt{tr}(\mathbf{G}^r(t)) = \mathtt{tr}([\dot{\mathbf{Y}}^r(t)][\dot{\mathbf{Y}}^r(t)]^\mathtt{T}). \tag{3.1}$$

The total energy signal $E^r(t)$ is an important property of the trial because it summarizes the magnitude of motion during a trial and condition. Our goal is to estimate time warping functions $h^r(t)$ such that the phase variation of $t \to \mathbf{Y}^r(h^r(t))$ is minimized across trials. One of the benefits of this signal is that it is invariant under rotations of the 3-dimensional variables.

The MINEIG criterion. We estimate the time warping functions for each trial or replication by minimizing the `MINEIG` criterion iteratively as formally explained in Ramsay and Silverman (2005) and in the Supplementary Material Section

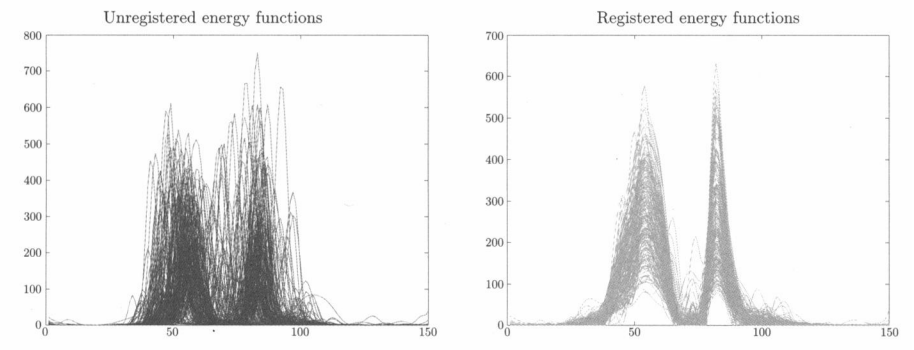

Figure 1. Example of energy profiles raw and aligned (small cone, 45° adduction). On the x-axis we show time and on the y-axis the value of the energy functions.

(S4.1). The main idea is to take one replication at a time, and to iteratively construct its warping function such that the shape of the warped energy is close to the shape of the mean energy across replications of the same condition. This procedure is done by iteratively optimizing the regularized objective function written as (S4.2) and results in a warping function per replication that minimizes the overall discrepancy of the energy curves for all replications from the mean energy function. The warping functions are strictly monotone smooth functions represented through a monotone smoother ($\mathcal{B}$-spline basis) as in Equation (S4.3).

In Figure 1 we show raw energy profiles and their aligned versions.

Recovering the aligned kinematic curves. Having estimated $h^r(t)$, the aligned velocity curves are $\dot{\mathbf{Y}}^r(h^r(t))$, while the positional curves can be obtained by integration:

$$\mathbf{Y}_0 + \int_0^t \dot{\mathbf{Y}}^r(h^r(u))du, \tag{3.2}$$

where $\mathbf{Y}_0$ corresponds to the initial hand configuration in the positional space.

4. Simulation Studies

We generated $p = 50$-dimensional data from a latent process of dimension $d_{true} = 4$. The dimensionality p of the observed simulated data roughly corresponds to the grasping data analyzed in further sections. We considered $T = 51$ time points and set $\mu \in \mathbb{R}^{p\times T}$ deterministically as a sinusoidal function of time with different amplitudes per coordinate: $\mu^k(t) = \sin\left([2 \cdot k/p] \cdot (t - k)\right), \quad k = 1, \ldots, p$; the entries of $\mathbf{B} \in \mathbb{R}^{p\times d}$ were drawn iid from $\mathcal{U}(0, 1)$; we set $\rho = 0.25$, and assumed $\Sigma(i, j) = \exp\left(-(i - j)^2/\theta_\Sigma\right)$ with $\theta_\Sigma = 0.01$.

Measures of goodness of fit. We summarize the mean square error (MSE) of observation r at a particular time slice t as $e_r(t) = (1/p)\sum_{i=1}^{p}(Y_i^r(t) -$

$\hat{Y}_i^r(t))^2$, and the mean integrated square error (MISE) of observation r along time as $\mathsf{e}_r = (1/T)\sum_{t=1}^{T} e_r(t)$. These statistics summarize the reconstruction error. Also, in a specific simulation S we obtain the mean error in the simulation as $\mathcal{E}_S = (1/R_S)\sum_{r=1}^{R_S} \mathsf{e}_r$ where R_S is the number of observations in simulation S. We report the average of $\mathcal{E}_S$ across independent simulations S, and we also report its corresponding standard error.

Number of required training samples. We explore the question of how many training examples are required to achieve a certain performance and we compare the differences between modelling μ free (unconstrained) and with splines with initialization through down-projection, keeping Σ free (unconstrained). We kept the latent dimension d fixed at $d_{true} = 4$, we generated a single test set of size 500, and ten different training sets for each specific number of training examples. Figure 2 shows that, in every case, performance improved in terms of reconstruction error as the amount of training data increased. We also notice that after 40 samples the performance levels off. There is no significant difference between modelling μ unconstrained (as a matrix in $\mathbb{R}^{p\times T}$) and μ with splines but, considering the number of parameters to be estimated, the best performance is achieved with the modelling of μ as splines.

In our simulation we considered $d_{true} = 4$. If we underestimate the true latent dimension d, then no matter how large the training sample, we would not be able to achieve the minimum Mean Integrated Squared Error (or the maximum loglikelihood). On the other hand, if we overestimate the latent dimension we can potentially achieve the optima due to the models being 'nested' as d increases, but we will need more samples. In the presented simulations we do not provide a quantification of how many more samples.

Latent dimension and reconstruction error. In the second simulation we study the behavior of different models when varying the value of the latent dimension. We considered modelling μ as a *constant* across dimensions but varying along time, as unconstrained or *free* (as a matrix in $\mathbb{R}^{p\times T}$), and with $B-$splines. We also considered two types of initialization: the matrix-normal MLE and the down-projection of a solution from a higher dimension. We investigated the performance in terms of reconstruction error, and we studied whether the model and learning procedure were able to recover the true dimensionality of the data. We set apart a single test set of 500 samples and 10 training sets of size 20 for each value of d. We considered a training set of size 20 because this number corresponds to the number of samples of a specific condition in a session.

In Figure 3 we display the average MISE on the test set, and the Bayesian Information Criterion (BIC) on the test set for each of the considered models, learning settings, and for various values of the latent dimension.

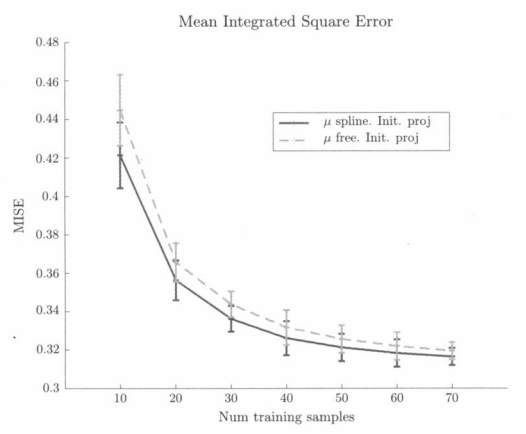

Figure 2. MGPFM: Average MISE in the test set as a function of number of training samples. The figure shows the results on a single test set of 500 samples using ten independent training sets of size 80. We report the average mean error in the test set for the ten simulations ($(1/10)\sum_{i=1}^{10}\mathcal{E}_{S_i}$) and its standard error. The dimensionality of the observed data was $p = 50$ and the latent dimension $d_{true} = 4$. We modelled Σ as unconstrained, the μ with splines and unconstrained, and initialized the learning algorithm with the down-projection.

In terms of MISE, modelling μ as a constant resulted in the worst performance. Modelling μ free yielded much better results in both initialization regimes. While modelling μ unrestricted works well in terms of the mean integrated square error, using splines drastically reduces the number of parameters being estimated and, everything else constant, should be preferred. In addition, using splines ensures that the μ is smooth. In terms of MISE, the initialization regime played a bigger role when using splines than in the unconstrained setting suggesting that in the more constrained case the algorithm is more susceptible to local optima and requires smarter initialization.

The true latent dimension was recovered through use of BIC whenever the learning method was initialized through the projecting procedure, and sometimes with the MLE initialization. In all cases, the BIC was characterized by a very fast drop until reaching the true value of the latent dimension d, and the steep decrease was followed by either a slower decrease or slight increase in the BIC. The clearest case occured when modelling μ with splines and performing initialization through the projection procedure. In the supplementary material we show how the model works in one run of the simulation.

5. Data Analysis

We analyzed the 48-dimensional finger motion captured data from 23 sessions

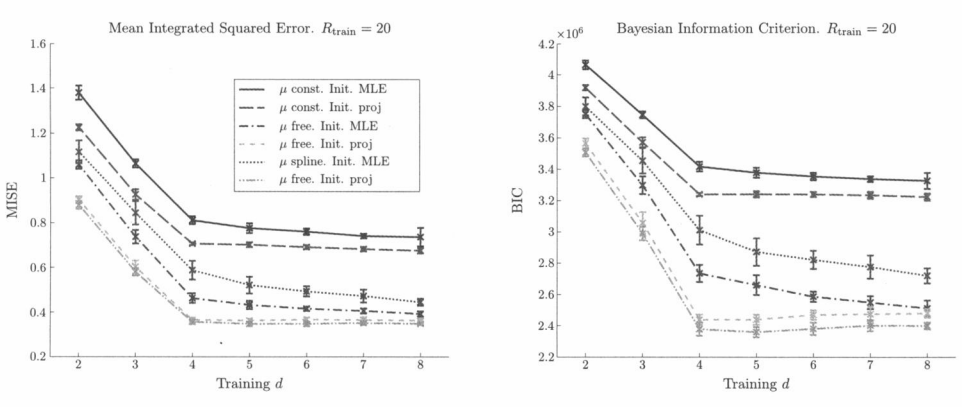

Figure 3. MGPFM: Average MISE in the test set, and BIC for different values of the latent dimensionality when varying the initialization regimes and the ways of modelling μ. The true latent dimension $d = 4$ is most obviously recovered when μ is modelled with splines and the initialization is through the projection procedure. When μ is modelled with splines the initialization regime has a large impact on performance.

of a rhesus monkey performing the reach-and-grasp task explained in Section 1. For the analysis we considered five conditions: small cone and small handle presented in positions 45° of flexion and 45° of adduction and for the small cone at 45° of abduction. In each condition we considered all trials across sessions, totalling in average 155 trials per condition.

We performed the analysis on the 3D velocity profiles, because velocity and speed of hand movements are thought to be encoded in the activity of single motor cortical neurons Moran and Schwartz (1999). Note that performing similar analysis with joint angle velocity is also possible Vinjamuri et al. (2007, 2010a,b).

We denote the observed velocity profiles as $\dot{\mathbf{Y}}^r(t)$ and fit the following model where $\boldsymbol{\mu}$ is modelled through splines,

$$\dot{\mathbf{Y}}^r(t) = \boldsymbol{\mu}^{\mathsf{S}}(t) + \mathbf{B}\dot{\mathbf{X}}^r(t) + \boldsymbol{\epsilon}^r(t). \tag{5.1}$$

Each reach-and-grasp replication lasted an average of 1.13 seconds, but each trial was of different length. In order to make it comparable, we smoothed data with a B-spline basis, resampling all trials onto 150 time slices.

For outlier removal we summarized each trial with its *energy function* (Equation 3.1) and clustered trials of a specific condition using the same function via $k-$means. We applied the clustering algorithm for several values of k (2, 3, and 4) and aggregated the resulting clusters removing the smallest group that contained at most 10% of trials and whose removal yielded the most visually uniform set of energy profiles. The MGPFM was applied to the preprocessed raw data and to the aligned data (as explained in Section 3). Figure 1 in Section 3 shows an example of the same data in these two states.

Analysis of performance varying models, alignment and latent dimensionality. We compared the performance of the MGPFM to two different baselines in terms of the MISE. We first compared our model against a *simple* baseline, namely modelling the data as a time varying mean; and secondly, we compared MGPFM against PCA, the prevailing approach in literature. In addition, we investigated whether aligning the data had an impact on the performance of the models in terms of MISE. Finally, we investigated the impact of varying the size of the latent dimension. We modelled μ with B-splines, Σ constrained, and we initialized the model with matrix normal MLE.

To apply PCA, we stacked the training trajectories into a 2D matrix of size $(R \cdot T) \times p$, R the number of training samples, $T = 150$, and $p = 48$ the number of kinematic variables. In applying this methodology we disregarded time correlations (as is usual in conventional PCA), and for this reason it does not make sense to align data before applying PCA.

We considered different conditions: each condition defined by the object and the orientation in which it was presented. We obtained the 10-fold cross validated MISE for the preprocessed raw data modelled with the mean, with PCA, and with the MGPFM. We did the same with the aligned data modelled with the mean and with the MGPFM.

Figure 4 shows a sample of the results of this experiment. Regardless of alignment, PCA or the MGPFM considerably outperform the mean. However the alignment helped in every setting including the mean, PCA, and MGPFM, and it gave a significant improvement in MISE, particularly when modelling the data only with the mean. Finally, MISE decreased in every case as the latent dimensionality increased. We notice that the impact that alignment has on the reduction of MISE is greater when the latent dimensionality is lower.

Results of a specific condition. We now discuss in some detail the results of estimating the model parameters for the small cone presented with 45° of abduction. We used the pre-processed and aligned data from all sessions (165 trials) and present the results of one of the ten folds (with 149 replications for training, and 16 for test). We modelled $\boldsymbol{\mu}$ with B-splines, Σ constrained, and we initialized the model with the matrix normal MLE. For simplicity, we set the latent dimensionality to be two.

In Figure 5 we plot the error in two ways: first, as a function of time, and second, integrated across time for each replication. We compare the MSE of MGPFM against a baseline of modelling only the mean. The top two plots show that there is much more variation during certain time periods in the trials and that the baseline is unable to capture this variation. The MGPFM significantly reduces the error (by approximately an order of magnitude) in those time periods by capturing the variation between trials.

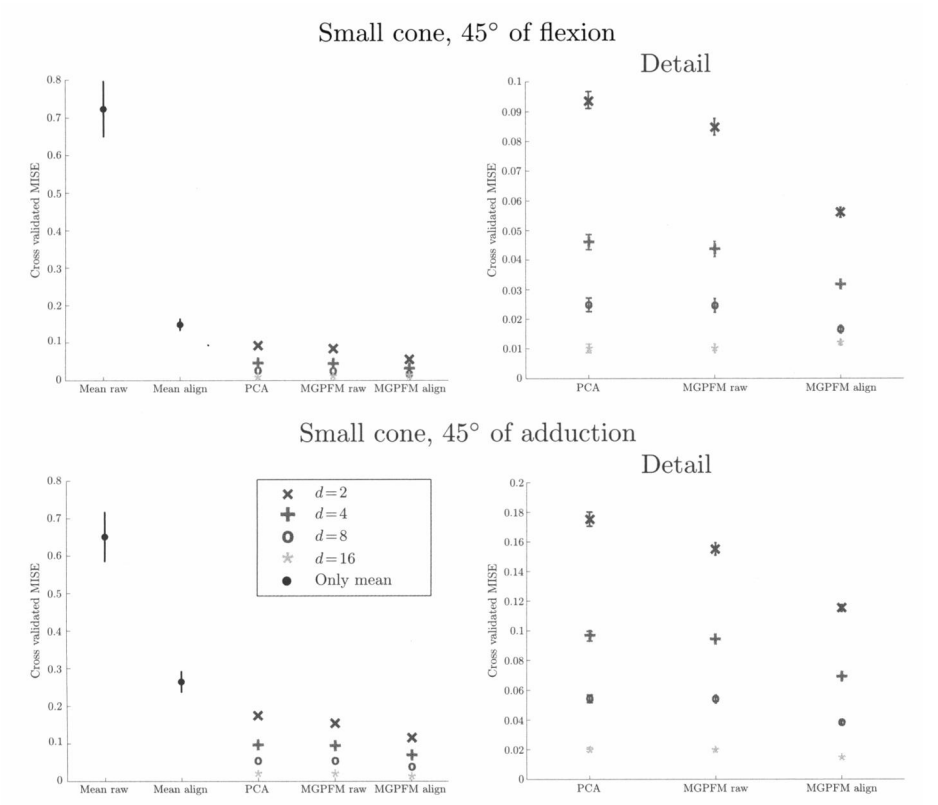

Figure 4. MGPFM: we show 10-fold cross validated MISE in grasping data for two conditions (small cone flexion and adduction), comparing the baseline model considering only the mean, PCA, MGPFM on raw pre-processed data and MGPFM on aligned data. Experiments were run for various sizes of latent dimensionality d. The MGPFM was applied modelling $\boldsymbol{\mu}$ with splines, Σ constrained, initializing with the MLE of the matrix normal distribution, and 50 iterations of the learning procedure. Observe that the MGPFM applied on aligned data achieves better results than other methods, but its advantage decreases as the size of the latent dimension increases.

In Figure 3 of the Supplementary Material we show the observed data, the model estimates and the residuals decomposed per marker and finger for a specific replication. These plots, which are representative of the grasping behavior in the dataset, show that the thumb's amount of movement is very small as compared to the amount of movement by all the other fingers. In other conditions (like the small handle) this is also the case, but the contrast is particularly prominent with the middle, ring, and index fingers. The MGPFM captures most of the variation leaving residuals close to zero.

Interpretation of learned parameters. One of the main features of the

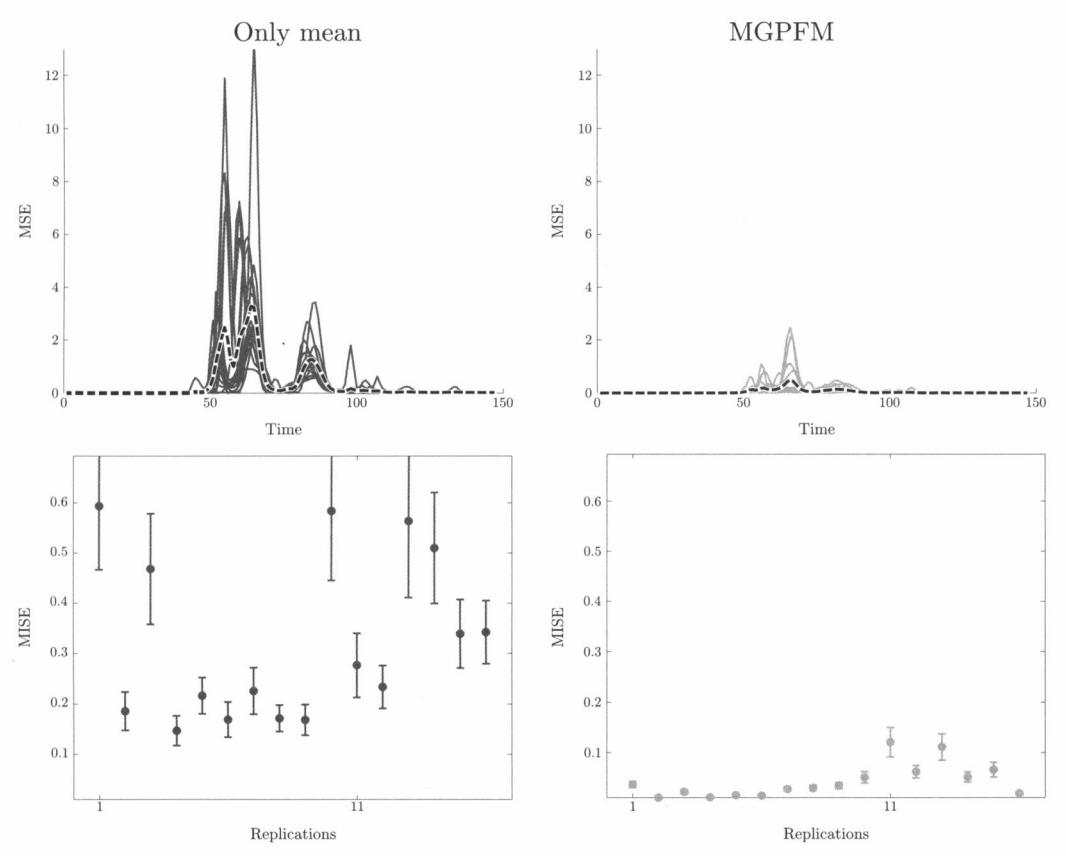

Figure 5. Error profiles in data (small cone, 45° abduction). In the upper panel, each line corresponds to a replication and the dashed line is the mean value. The lower panel displays the mean integrated square error per replication. The MGPFM reduces the error significantly, as compared to the baseline of modelling only the mean (by approximately an order of magnitude).

MGPFM is that its parameters can be interpreted. Parameter $\boldsymbol{\mu}$ is a trajectory in the velocity space and, through (3.2), we are able to obtain corresponding postures in the position space. In Figure 6 we show time slices of $\boldsymbol{\mu}$ projected onto the hand space that summarize the main features of the $\boldsymbol{\mu}$ trajectory. The mean parameter $\boldsymbol{\mu}$ captures the shared behavior across all trials. In this particular condition, this behavior consists of five *epochs* that correspond to the hand starting in a neutral position, followed by a slight opening of the fingers in a synchronized fashion, back to a neutral position, after which the subject spreads his fingers slightly before going back to neutral position. All trials in this condition showed this pattern.

In contrast to the parameter $\boldsymbol{\mu}$ that encodes behavior shared among all

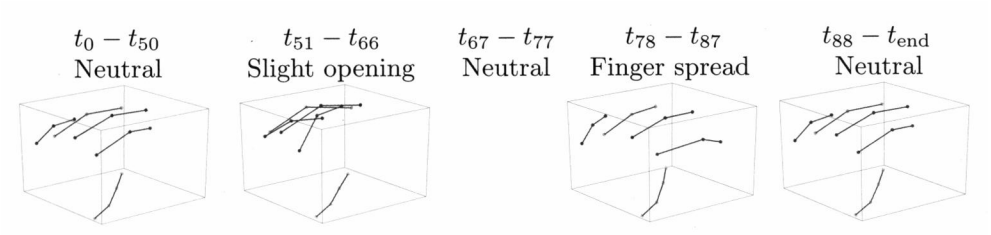

Figure 6. The trajectory $\hat{\boldsymbol{\mu}}$ (projected onto the position space) represents what is shared by all replications of a condition. Here we show the visualization of $\hat{\boldsymbol{\mu}}$ for the small cone at 45° of abduction. The trajectory presents five *epochs* corresponding to hand postures: all plots are in the same scale and the third epoch looks very similar to the first and last epochs (neutral position), so we omit it. The trajectory begins at a neutral position, followed by a slight opening of the grasp through a synchronized movement of fingers and a slight rotation, then back to the neutral hand configuration after which the fingers spread slightly (after the subject releases the object) and back to a neutral position. All replications for this condition followed this pattern; they differentiated among themselves with the movement modelled through the loading matrix and the MGP term.

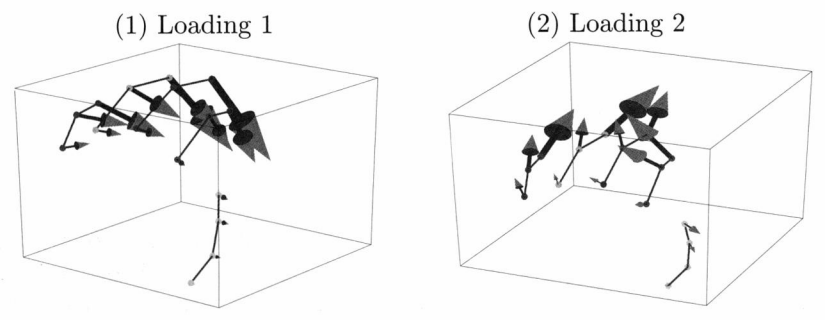

Figure 7. Visualization of loadings encoded in $\hat{\mathbf{B}}$ for the small cone presented at 45° abduction. The first loading corresponds to synchronized opening-closing of the hand; the second loading to curling of the fingers wrapping around the cone. The estimation of $\hat{\boldsymbol{B}}$ is explained in the Supplementary Material Section S1.2 and it can be reduced to solving a specific linear regression problem (Equation (S1.9)).

replications, the term that includes the latent factors $\dot{X}$ and the loading matrix $\mathbf{B}$ corresponds to the ways in which a replication differentiates itself from the other replications. The factors $\dot{X}$ encode what is specific to each trial. Figure 8 shows $\dot{X}$ for two example replications superposed on the distribution of factors for all replications. The factors for these two representative replications differ.

The estimate of the loading matrix $\hat{\mathbf{B}}$ is defined in Equations S2.8–S2.11 of the Supplementary Material, and even though $\hat{\mathbf{B}}$ is only identifiable up to rotation we are able provide a limited interpretation after sorting the columns of

$\hat{\mathbf{B}}$ in decreasing order of their norm. Figure 7 shows both columns of $\hat{\mathbf{B}}$ as a set of $K = p/3 = 16$ vectors that correspond to the direction and relative magnitude of change for each of the K markers. Column i of $\hat{\mathbf{B}}$, denoted by $\hat{b}_{i,\cdot} \in \mathbb{R}^p$, is constructed by stacking K 3D vectors, each of which can be interpreted as the 3D direction of change of a particular marker as the learned latent variable $\dot{X}(t)$ changes in the direction of the ith dimension of the low dimensional space. Each arrow in Figure 7 corresponds to a marker and is determined by $Y_0 + \alpha \cdot \hat{b}_{i,\cdot}$ where $Y_0 \in \mathbb{R}^p$ is the mean hand postural configuration at $t = 1$ of the training set and $\alpha \in \mathbb{R}$ is a scalar that controls the size of the arrow. Y_0 is used simply to project the data (which was learned in the velocity space) into the positional space, and its selection can be arbitrary; multiplying by α preserves the relative contribution to the motion of a particular marker with respect to the other markers.

The first loading, or first column, of $\mathbf{B}$ (left panel) encodes a synchronized closing-opening motion of the fingers, whereas the second loading, or second column, of $\mathbf{B}$ (right panel) encodes a movement that happens at a somewhat different rate and direction per finger and whose net effect captures a wrapping (or curling) of the fingers (as if around the cone). These two loadings represent the ways in which trials differ amongst themselves in the magnitude of the grasp opening and the emphasis of the curling motion.

In this analysis we considered $d = 2$ for illustrative purposes (in the simulations we showed how we can use BIC to select the dimension). In principle though, there is nothing that prevents us from providing an interpretation of the learned loadings when $d > 2$ in the same manner as we did above. In fact, columns of the loading matrix with smaller norm correspond to smaller variability of the kinematics of the hand. However, as d increases, the loadings will most likely be harder to interpret due to the fact that we start modelling noise (as in PCA).

Under the conditions specified in Section 2.2, the factor loadings are identifiable. In this case, it makes sense to interpret the learned latent factors in the model. Whereas $\hat{\dot{X}}$ is estimated in the velocity space, it is more intuitive to visualize the differences between trials on the position space by integrating the latent factors along time and adding the corresponding initial hand posture (as in (3.2)). In this way we are able to compare the two replications in the positional space at specific time periods, for instance, between time points 50 and 58 (it is valid to compare the two replications at the same time period because, through the alignment procedure, we have accounted for the phase variation). We observe that, whereas the first factor corresponding to replication 1 transitions from $\hat{\dot{X}}_1(t = 50) = +54.19$ to $\hat{\dot{X}}_1(t = 58) = -684.8$ (with a net change of -738.99), the first factor of replication 2 changes from $\hat{\dot{X}}_1(t = 50) = +73.21$ to

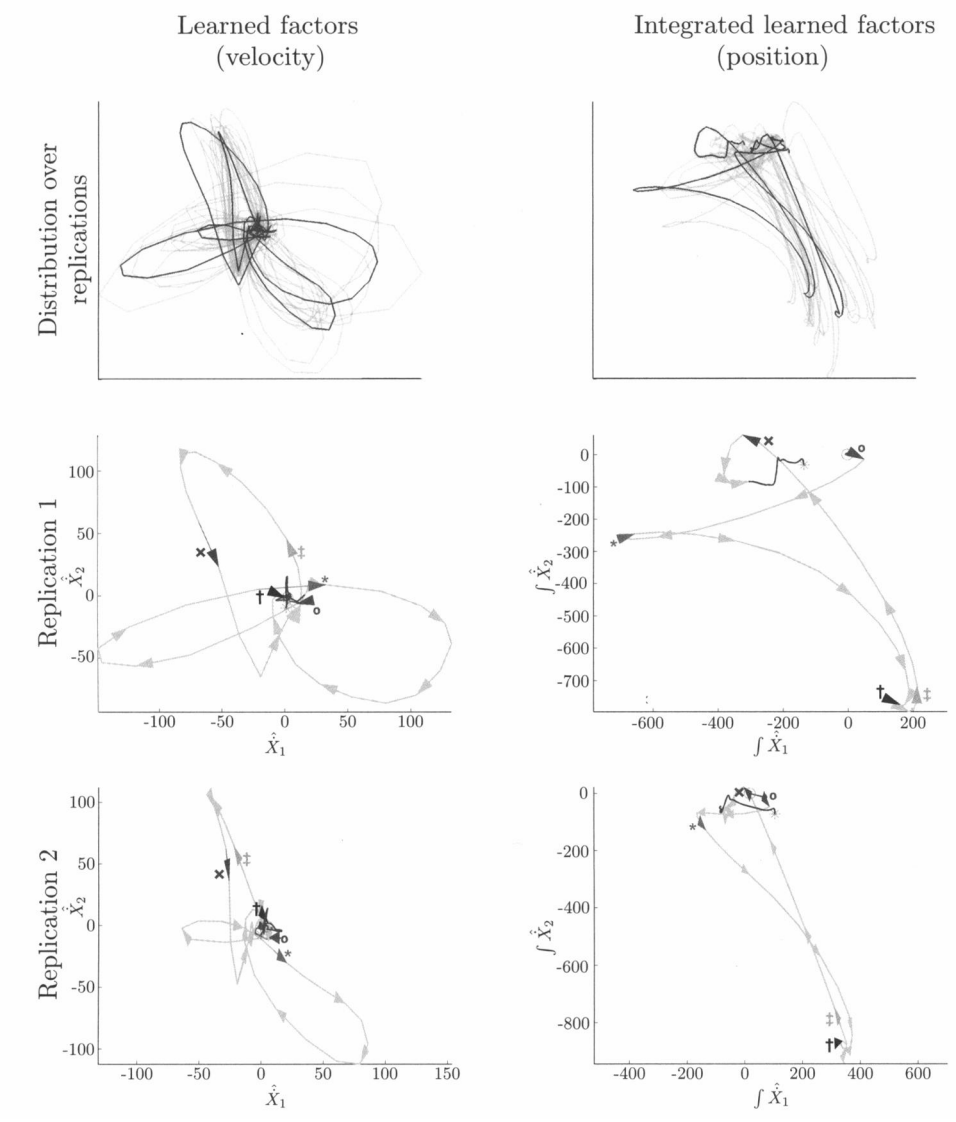

Figure 8. Learned factors $\hat{\hat{X}}$ for condition: small cone, 45° abduction. In the top panel we show (in green) the distribution of learned factors in the velocity space (left) and their integrated version on the positional space (right). This figure depicts differences between trials in the space of learned factors. On this plot we overlap two exemplary trials. In the middle and lower panel we show details of these replications: the shape and values they display are different. The starting point of the trial is denoted by an open circle, the end position, by a star. There are arrows along the trajectory show the direction of movement. Arrows marked with different symbols represent time and allow for comparison between trajectories: arrow with circle (33%), arrow with star (40%), arrow with spade (50%), arrow with double spade (54%), arrow with cross (60%). In Figure 9 we show how difference between the integrated learned factors in these two trials manifest on hand posture.

$\hat{\hat{X}}_1(t = 58) = -178.5$ (with a net change of -251.71). While the net change is not meaningful by itself, the relative change is. The first/corresponding column in the loading matrix $\mathbf{B}$ suggests that these changes should result in an exaggerated opening of the hand in replication 1 as compared to replication 2. And, indeed by visualizing $\hat{Y}$ (and the observed data Y) we verify that the hand in replication 1 opens further than in replication 2 (see Figure 9).

Thus each of the elements of the MGPFM can be mapped to the actual configuration of the grasping curves and we can provide physical and intuitive interpretation, albeit somewhat limited by the issue of identifiability up to rotation. Furthermore, we can differentiate the variability that corresponds to a specific replication from the variability shared among all trials from a specific condition. Finally, we are also able to accurately recover (in terms of reconstruction error) hand configurations from the estimated parameters and factors.

6. Discussion

In this paper we formulated a dynamic factor model based on Multivariate Gaussian Processes to study grasping kinematics. We developed an algorithm for inference and parameter estimation for the MGPFM and discussed variations of the model and algorithm. We showed in simulations that our model outperforms PCA in the reconstruction of error when alignment is applied. In contrast with PCA or SVD, we are able to differentiate sources of variation that can potentially be used to design robotic primitives for grasping devices. Ciocarlie, Goldfeder, and Allen (2007), for example, show how to use PCA for such purposes; in contrast, our MGPFM incorporates time modelling into the primitives. The MGPFM can also be extended by assuming prior distributions in the parameters (for instance, in the loading matrix), and can capture long range correlations that can potentially improve the prediction of coordinated dexterous hand motions. The MGPFM is also easy to adapt to new settings — we can add sensors, change the representation to joint angles, and the same algorithms apply in principle. Furthermore, though we have not addressed the application here, the MGPFM can potentially be extended to incorporate neural data as a controller for the kinematic motion of a robotic arm. Saleh, Takahashi, and Hatsopoulos (2012), for example, decode from neural data PCA-reduced kinematic configurations in a short period of time; a potential extension of our model would explicitly model and exploit temporal structure.

The MGPFM probabilistically models the relevant grasping structure and separates it from noise; but our core methodological contribution is a strategy to decompose and reduce the dimensionality of the variation of the data according to the experimental structure (time, condition and replications). The decomposition of variance in the grasping datasets relied on the application of a

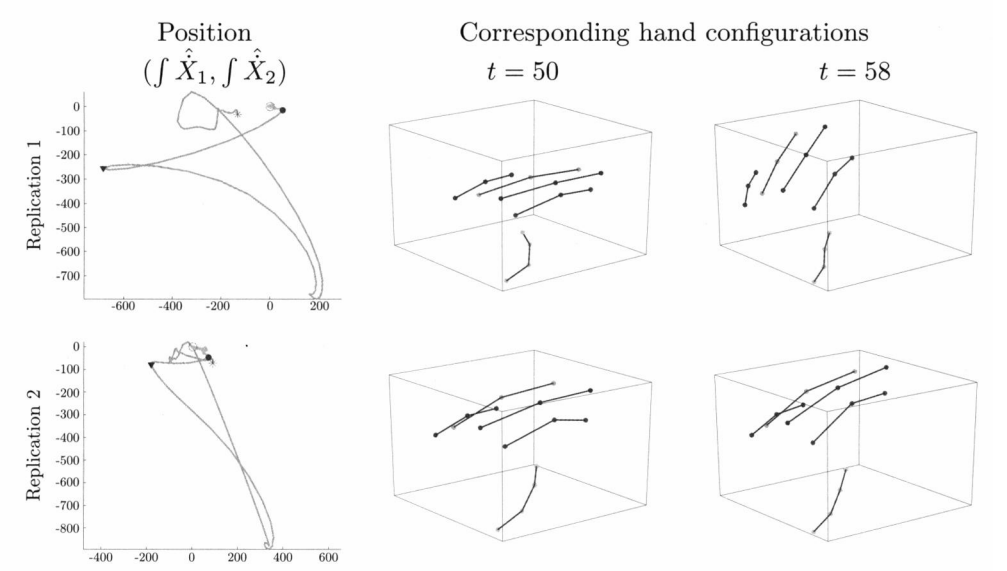

Figure 9. Interpretation of latent factors showing differences between replications. On the left we plot $(\int \hat{X}_1(t), \int \hat{X}_2(t))$ as a function of t. The start of the trial is at the open circle, the solid dot corresponds to $t = 50$, the triangle to $t = 58$ and the star to the end of the trial. Middle and right panels: hand configurations corresponding to those time points. The interaction between the first latent factor (moving negatively) and the corresponding loading (Figure 9 panel 1) corresponds to an opening of the fingers in a synchronized manner – this movement differs between the two replications and leads to an exaggerated opening of the hand in replication 1 (top panel).

multivariate functional alignment procedure. A major product of this approach is the decomposition of variability between what is common in replications and what is specific for each trial; it also provides clear interpretation in the space of grasp postures. In particular, visualizations of the shared mean trajectory $\boldsymbol{\mu}$, of the axis of variation in replications encoded in the loading matrix $\mathbf{B}$, and of the specific differences in particular trials summarized in the latent factors $\mathbf{X}$ helped to explain variability in grasping movements.

Acknowledgements

This work was supported by grants RO1 MH064537 (L. Castellanos and R. Kass) and R90 DA023426 (L. Castellanos). V. Q. Vu was supported in part by NSF Postdoctoral Fellowship DMS-09-03120. The work of S. Perel was supported by the Defense Advanced Research Projects Agency (DARPA) and SPAWAR System Center Pacific (SSC Pacific) under Contract No. N66001-12-C-4027. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of DARPA and

SSC Pacific. We thank J. Huang for helpful discussions regarding analyses and proof reading.

References

Ciocarlie, M., Goldfeder, C. and Allen, P. (2007). Dimensionality reduction for hand-independent dexterous robotic grasping. *IROS*, 3270-3275.

d'Avella, A. and Bizzi, E. (2005). Shared and specific muscle synergies in natural motor behaviors. *PNAS* **102**, 3076-3081.

Dawid, A. P. (1981). Some matrix-variate distribution theory: notational considerations and a bayesian application. *Biometrika* **68**, 265-274.

Dutilleul, P. (1999). The mle algorithm for the matrix normal distribution. *J. Statist. Comput. Simulation* **64**, 105-123.

Fox, E. and Dunson, D. (2011). Bayesian nonparametric covariance regression. arXiv preprint arXiv:1101.2017.

Fyshe, A., Fox, E., Dunson, D. and Mitchell, T. (2012). Hierarchical latent dictionaries for models of brain activation. In *AISTATS*, April 2012.

Mason, C. R., Gomez, J. E. and Ebner, T. J. (2001). Hand synergies during reach-to-grasp. *J. Neurophysiology* **86** 2896-2910.

Mason, C. R., Theverapperuma, L. S., Hendrix, C. M. and Ebner, T. J. (2004). Monkey hand postural synergies during reach-to-grasp in the absence of vision of the hand and object. *J. Neurophysiology* **91**, 2826-2837.

Moran, D. W. and Schwartz, A. B. (1999). Motor cortical representation of speed and direction during reaching. *J. Neurophysiology* **82**, 2676-2692.

Pesyna, C., Pundi, K. and Flanders, M., Coordination of hand shape. *J. Neuroscience* **31**, 3757-3765.

Ramsay, J. and Silverman, B. W. (2005). *Functional Data Analysis.* Springer Series in Statistics. Springer.

Ramsay, J., Hooker, G. and Graves, S. (2009). *Functional Data Analysis with R and MATLAB.* Springer.

Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian Processes for Machine Learning.* Adaptive Computation And Machine Learning. Mit Press.

Saleh, M.; Takahashi, K. and Hatsopoulos, N. G. (2012). Encoding of coordinated reach and grasp trajectories in primary motor cortex. *J. Neuroscience* **32**, 1220-1232.

Santello, M.; Flanders, M. and Soechting, J. F. (1998). Postural hand synergies for tool use. *J. Neuroscience* **18**, 10105-10115.

Soechting, J. F. and Flanders, M. (1997). Flexibility and repeatability of finger movements during typing: analysis of multiple degrees of freedom. *J. Comput. Neuroscience* **4**, 29-46.

Thakur, P. H.; Bastian, A. J. and Hsiao, S. S. (2008). Multidigit movement synergies of the human hand in an unconstrained haptic exploration task. *J. Neuroscience* **28**, 1271-1281.

Todorov, E. and Ghahramani, Z. (2004). Analysis of the synergies underlying complex hand manipulation. In *EMBS*, 4637-4640.

Vinjamuri, R., Mao, Z.-H., Sclabassi, R. and Sun, M. (2007). Time-varying synergies in velocity profiles of finger joints of the hand during reach and grasp. In *EMBS*, 4846-4849. IEEE.

Vinjamuri, R., Sun, M., Chang, C.-C., Lee, H.-N., Sclabassi, R. J. and Mao, Z.-H. (2010a). Temporal postural synergies of the hand in rapid grasping tasks. *IEEE Transactions on Information Technology in Biomedicine* **14**, 986-994.

Vinjamuri, R., Sun, M., Chang, C.-C., Lee, H.-N., Sclabassi, R. J. and Mao, Z.-H. (2010b). Dimensionality reduction in control and coordination of the human hand. *IEEE Transactions on Biomedical Engineering* **57**, 284-295.

Wang, J. M., Fleet, D. J. and Hertzmann, A. (2008). Gaussian process dynamical models for human motion. *IEEE Transactions Pattern Analysis Machine Intelligence* **30**, 283-298.

Yu, B. M., Cunningham, J. P., Santhanam, G., Ryu, S. I., Shenoy, K. V. and Sahani, M. (2009). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *J. Neurophysiology* **102**, 614-635.

Machine Learning Department, Carnegie Mellon University, Gates Hillman Center 8010, 5000 Forbes Avenue, Pittsburgh, PA 15213-3891, USA.

E-mail: lucia.castellanos.pb@gmail.com

Department of Statistics, The Ohio State University, Columbus, OH 43210-1247, USA.

E-mail: vqv@stat.osu.edu

Department of Biomedical Engineering and the Center for Neural Basis of Cognition, Carnegie-Mellon University, Pittsburgh, PA 15213-3891, USA.

E-mail: sagi@cmu.edu

Department of Neurobiology, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15213-2536, USA.

E-mail: abs21@pitt.edu

Department of Statistics, Carnegie-Mellon University, Pittsburgh, PA 15213-3891, USA.

E-mail: kass@stat.cmu.edu

(Received August 2013; accepted May 2014)