Selecting Sensitive Parameter Subsets in Dynamical Models With Application to Biomechanical System Identification

Ahmed Ramadan

Mem. ASME
Department of Mechanical Engineering,
MSU Center for Orthopedic Research (MSUCOR),
Michigan State University,
428 S. Shaw Ln,
East Lansing, MI 48824
e-mail: ramadana@msu.edu

Connor Boss

Mem. ASME
Department of Electrical and Computer Engineering,
MSU Center for Orthopedic Research (MSUCOR),
Michigan State University,
East Lansing, MI 48824
e-mail: bossconn@egr.msu.edu

Jongeeun Choi¹

Mem. ASME
School of Mechanical Engineering,
Yonsei University,
Seoul 03722, Republic of Korea
e-mail: jongeeunchoi@yonsei.ac.kr

N. Peter Reeves

Sumaq Life LLC,
East Lansing, MI 48823;
Department of Osteopathic Surgical Specialties,
MSU Center for Orthopedic Research (MSUCOR),
East Lansing, MI 48824
e-mail: reevesn@icloud.com

Jacek Cholewicki

Department of Osteopathic Surgical Specialties,
MSU Center for Orthopedic Research (MSUCOR),
Michigan State University,
East Lansing, MI 48824
e-mail: cholewic@msu.edu

John M. Popovich, Jr.

Department of Osteopathic Surgical Specialties,
MSU Center for Orthopedic Research (MSUCOR),
Michigan State University,
East Lansing, MI 48824
e-mail: popovi16@msu.edu

Clark J. Radcliffe

Fellow ASME
Department of Mechanical Engineering,
MSU Center for Orthopedic Research (MSUCOR),
Michigan State University,
East Lansing, MI 48824
e-mail: radcliffe@egr.msu.edu

Estimating many parameters of biomechanical systems with limited data may achieve good fit but may also increase 95% confidence intervals in parameter estimates. This results in poor identifiability in the estimation problem. Therefore, we propose a novel method to select sensitive biomechanical model parameters that should be estimated, while fixing the remaining parameters to values obtained from preliminary estimation. Our method relies on identifying the parameters to which the measurement output is most sensitive. The proposed method is based on the Fisher information matrix (FIM). It was compared against the nonlinear least absolute shrinkage and selection operator (LASSO) method to guide modelers on the pros and cons of our FIM method. We present an application identifying a biomechanical parametric model of a head position-tracking task for ten human subjects. Using measured data, our method (1) reduced model complexity by only requiring five out of twelve parameters to be estimated, (2) significantly reduced parameter 95% confidence intervals by up to 89% of the original confidence interval, (3) maintained goodness of fit measured by variance accounted for (VAF) at 82%, (4) reduced computation time, where our FIM method was 164 times faster than the LASSO method, and (5) selected similar sensitive parameters to the LASSO method, where three out of five selected sensitive parameters were shared by FIM and LASSO methods. [DOI: 10.1115/1.4039677]

Keywords: parameter estimation, Fisher information, LASSO, sensitive parameters, head position tracking

1 Introduction

The estimated parameters in models of sensorimotor control provide insight into the relative contributions of various feedback pathways for control of several body segments such as head-neck [1–4], trunk [5–7], and arm [8–10] models. However, some difficulties are associated with these models, which limit their application. Difficulties occur mainly because these models can produce a large number of parameters that need to be estimated [11]. At the same time, the information in experimental data is limited because subjects do not tolerate a long data collection process before fatiguing, and noninvasive external sensors are not available to measure all signals of interest.

Overfitting is a consequence of estimating too many parameters in the model with limited measurements. Overfitting is well recognized in polynomial fitting, e.g., fitting a 14th order polynomial to 15 data points yields a perfect fit but no useful information about the relevance of fitted model parameters to the behavior tested. Similarly, as parametric models become more complex with more parameters to be estimated, goodness of fit may improve. However, the estimated parameters may have more variability [7,12], because multiple parameter solutions yield very similar model responses [1]. This difficulty leads to poor identifiability in the estimation problem [13,14]. Moreover, with increased model complexity, the estimation is more likely to be misled by noise possibly resulting in incorrect parameter estimates [15]. Goodness of fit must be balanced against the number of model parameters estimated to avoid overfitting with too many parameters for the information in data.

A good trade-off between number of parameters estimated and goodness of fit can be achieved by fixing some parameters to values obtained from preliminary estimation while estimating others [7,12]. We propose only those sensitive parameters significantly

¹Corresponding author.

Manuscript received October 5, 2017; final manuscript received March 2, 2018; published online May 8, 2018. Assoc. Editor: Guy M. Genin.

Table 1 The subjects' demographic characteristics presented as mean $\pm$ standard deviation

	Females	Males
No. of subjects	7	3
Height (m)	1.66 ± 0.11	1.78 ± 0.07
Weight (kg)	55.3 ± 11.8	87.7 ± 17.7
Age (years)	22.3 ± 1.2	34.0 ± 11.8

affecting the model response be estimated. The selection of parameters sensitive to the data allows those parameters to be accurately estimated by that data.

Sensitive parameters have been discussed in the literature in the setting of optimal experiment design, e.g., experiments were designed such that the parameters would be identifiable from the experimental data [12,16,17]. In this study, we consider an inverse view where sensitive parameters are identified from the experimental data to provide an optimal trade-off between number of parameters estimated and goodness of fit. Our method uses Fisher information matrix (FIM) to select sensitive model parameters using available experimental data in a novel problem formulation for biomechanical modeling.

Other methods of achieving this trade-off have been reported in the literature. An ad hoc method of selecting a model out of a few possible combinations was proposed in Ref. [7], but the combinations were not exhaustive. Akaike's information criterion [18,19] and information complexity [20] are very expensive computationally because they require every combination of the model parameters being estimated to determine their goodness of fit and there are 4083 combinations of two parameters or more chosen from a 12-parameter model in our example's analysis. Nonlinear least absolute shrinkage and selection operator (LASSO) [21] also requires multiple estimations to update a tuning parameter. The

number of estimations in LASSO is, in general, smaller than the number of parameter combinations. Other machine learning methods provide black-box models that have no physiological interpretations, e.g., Gaussian process regression in gait kinematics [22].

The goal of this study is to introduce and evaluate a novel FIM method to select sensitive parameters in modeling sensorimotor control, in this case the head-neck segment. Our method improves on previously published methods, because (1) it is faster as it relies on a single preliminary estimation, and (2) the full parametric model has a physiological interpretation. Early work of our method was presented without evaluating its performance [23]. In this study, we demonstrate sensitive parameter selection is robust to the method used by comparing our FIM method to the LASSO method [21]. Moreover, we compare the two methods in terms of goodness of fit, parameter confidence intervals, and computation time. This information will guide modelers on the pros and cons of this method.

2 Methods

2.1 Subjects. Ten healthy subjects participated in the experimental study (Table 1). They did not have any history of neck pain lasting more than three days or any neurological motor control impairment. The Michigan State University's Biomedical and Health Institutional Review Board approved the test protocol. The subjects signed an informed consent before participating.

2.2 Data Collection. A head position tracking task in the horizontal, yaw, plane (Fig. 1(a)) provided data on subject head response to vision-based commands. The experiment was statistically shown to be reliable [24]. Each subject wore a helmet with two attached string potentiometers (SP2-50, Celesco, Chatsworth, CA) to measure the head rotation. On a monitor (SyncMaster SA650, Samsung; height 27 cm, width 47.5 cm) located 1 meter

(a)

(b)

Fig. 1 (a) Experimental setup for head position tracking. The reference marker is $r(t)$ and the head position marker is $y(t)$. (b) The block diagram of the head position tracking model adopted from Refs. [3] and [4]. The transfer functions come from Refs. [2] and [4] after accounting for the change in output from acceleration to position.

Fig. 2 The best-fit model output (during a trial) when estimating the FIM- and LASSO-selected sensitive parameters. The experimental (EXP) output is a subject's response.

away from the subject, two markers were displayed: reference command signal $r(t)$ and the measured head position response $y(t)$. Subjects were instructed to rotate their heads in the horizontal plane to track the reference signal. The input signal $r(t)$ was a pseudorandom sequence of steps with random step duration and amplitude, such that the amplitudes were bounded in the range ± 4 deg (see "Input" in Fig. 2). Each subject performed three 30 s trials. Data were collected using a data acquisition system (cDAQ-9172, National Instruments, Austin, TX) and sampled at 60 Hz.

2.3 Parametric Model. The block diagram of head position tracking model (Fig. 1(b)) adopted published physiological and feedback control models [2–4], where the output was modified to be angular position instead of angular acceleration. This change in output from acceleration to position does not change the behavior

modeled nor the definitions of the physiological parameters defined in those previous investigations.

The model parameters (physiological descriptions given in Table 2) are divided into two groups. One group has the candidate parameters, out of which we would select the sensitive parameters to be estimated. They are

$$\Theta := [K_{vis} \ K_{ver} \ K_{cer} \ \tau \ \tau_{1A} \ \tau_{CNS1} \ \tau_c \ \tau_{CNS2} \ \tau_{MS1} \ \tau_{MS2} \ B \ K]^T \quad (1)$$

The second group, $\Theta_{fixed} := [J \ T_c]^T$, was fixed to values found in the literature [2]. The bounds in Table 2 were obtained from preliminary parameter estimates discussed in Sec. 3.

2.4 Parameter Estimation. Parameter estimates define physiological models from measured task response. Parameter normalization is required because physiological parameters often have quite different magnitudes, which cause numerical problems in computer optimization. Each normalized parameter $\Phi(i)$ for a model parameter $\Theta(i)$ is given by

$$\Phi(i) = \frac{\Theta(i) - \mu(i)}{\sigma(i)} \in [-1, 1], \quad \forall i \in \{1, 2, \dots, n_\Theta\} \quad (2)$$

for the vector Θ of n_Θ parameters, μ is an offset to make the true bounds symmetric around 0, and σ is a gain that normalizes the symmetric parameter bounds to $[-1, 1]$. Estimates were computed by minimizing least squared error, i.e.

$$\hat{\Theta}_N = \underset{\Phi}{\operatorname{argmin}} \left(\frac{1}{N} \sum_{k=1}^N e^T(k, \Phi) e(k, \Phi) \right) \quad (3)$$

where $e(k, \Phi) = y(k) - \hat{y}(k, \Phi)$ is the error at data sample k , y is the measured head position output, $\hat{y}(k, \Phi)$ is the simulated model output computed at sample k and normalized parameter Φ , and N is the number of observation samples.

Goodness of fit of the simulated model data to the experimental data was measured by variance accounted for (VAF) [7]

Fig. 3 An outline of model-based bootstrap [25] using 1000 replicates to estimate 95% confidence interval (CI) of the parameter vector θ

Table 2 The neurophysiological parameters of the head position tracking task along with their physiological description. The lower and upper bounds were obtained from preliminary estimation of the subjects' data (Sec. 3).

Parameters		Max	Min	Description
Θ	K_{vis}	10^3	50	Visual feedback gain [4]
	K_{ver}	10^4	500	Vestibular feedback gain [2]
	K_{ccr}	300	1	Proprioceptive feedback gain [2]
	τ	0.4	0.1	Visual feedback delay [4]
	τ_{1A}	0.2	0.01	Lead time constant of the irregular vestibular afferent neurons [2]
	τ_{CNS1}	1	0.05	Lead time constant of the central nervous system [2]
	τ_c	5	0.1	Lag time constant of the irregular vestibular afferent neurons [2]
	τ_{CNS2}	60	5	Lag time constant of the central nervous system [2]
	τ_{MS1}	1	0.01	First lead time constant of the neck muscle spindle [2]
	τ_{MS2}	1	0.01	Second lead time constant of the neck muscle spindle [2]
	B	5	0.1	Intrinsic damping [2,4]
	K	5	0.1	Intrinsic stiffness [2,4]
Θ_{fixed}	J	0.0148	0.0148	Head inertia [2]
	T_c	0.1	0.1	Torque converter time constant [2]

$$VAF(\Phi) = \left[1 - \frac{\sum_{k=1}^N (y(k) - \hat{y}(k, \Phi))^2}{\sum_{k=1}^N (y(k))^2} \right] \times 100\%. \quad (4)$$

As VAF approaches 100%, the model response $\hat{y}(k, \Phi)$ approaches the measured response $y(k)$ and goodness of fit is better.

Parameter 95% confidence interval (CI) estimates are critical for evaluating the success of any method. We used model-based bootstrap [25] (Fig. 3). We generated 1000 bootstrap replicates of the experimental data $(y_1^*, \dots, y_{1000}^*)$ per subject. Then, we estimated the parameters θ for every replicate to get $(\hat{\theta}_1^*, \dots, \hat{\theta}_{1000}^*)$. The 95% CI was computed from 2.5 to 97.5 percentiles of the empirical distribution of $\hat{\theta}$. Paired t-tests with significance level set at $p \leq 0.05$ were used to statistically compare the results between the methods.

2.5 Sensitive Parameter Selection. Sensitive parameters were selected using FIM (Appendix A) that is a real $n_\theta \times n_\theta$ matrix denoted by $\mathbb{I}$, where n_θ is the number of model parameters. FIM is computed at the nominal values of the model parameters Φ_0 , which were obtained via preliminary estimation. The diagonal elements of $\mathbb{I}^{-1}$ are estimates of parameters' variance. The FIM method for selecting sensitive parameters (Appendix C) scans all possible parameter combinations to ensure that the estimation error variance ($\mathbb{I}^{-1}$) is below a user-defined threshold $\sigma_{allowable}^2$, and maximize the number of sensitive parameters (n_θ) to improve goodness of fit. The user must provide FIM $\mathbb{I}$ (Appendix A), allowable normalized variance $\sigma_{allowable}^2$, and the desired minimum number of sensitive parameters n_{min} .

Least absolute shrinkage and selection operator (Appendix B) is an alternative method that shrinks insensitive normalized parameters to zero while estimating the remaining sensitive parameters. This is done by adding a new term (regularization) to the cost function in Eq. (3). In our application, shrinking the normalized parameters to zero corresponds to shrinking the actual parameters to the values centered between their bounds, which is likely not as correct as typical physiological values. Therefore, a modification of the regularization in LASSO was made (Appendix C) to ensure that parameters were pushed toward their typical values [21].

3 Results

Nominal parameter values are required to compute FIM and to be the LASSO typical parameters. Adopting values from prior literature yielded a poor fit, which might be because the literature either used a simpler model [4] or did not have a visual tracking task [2]. Therefore, we did preliminary parameter estimation to obtain the nominal parameter values over all three trials ($i = 1, 2, 3$) for each subject. The cost function in Eq. (3) becomes

$$\frac{1}{3} \sum_{i=1}^3 \left(\frac{1}{N} \sum_{k=1}^N e_i^T(k, \Phi) e_i(k, \Phi) \right) \quad (5)$$

where i is the trial index, and e_i is the error of the i th trial. Minimizing this cost function yields the best-fit parameters of a subject for the three trials collectively (Table 3).

In the FIM method, we computed FIM at each subject's best-fit parameters to estimate parameters' variance for each subject. From each FIM, a subset or more of subject-specific sensitive parameters were selected. Then, only the common parameter subsets among all subjects were selected. This is a more accurate process than building one FIM from the mean of subjects' parameters and selecting sensitive parameters based on this averaged FIM. We applied the FIM method to the experimental data with ten equally spaced values of $\sigma_{allowable}^2$ from 0.05 to 0.95 and $n_{min} = 2$ (Fig. 4). The number of selected sensitive parameters n_θ^* was between 3 and 5. We selected $n_\theta^* = 5$ at $\sigma_{allowable}^2 = 0.55$, because this solution included a single subset $N_{candidate} = 1$. The selected subset was $\theta^{FIM} = \{K_{vis}, K_{ccr}, \tau, \tau_{1A}, \tau_{CNS1}\}$.

In the LASSO method, we considered the mean of nominal parameter estimates as the LASSO typical parameters. If we used subject-specific parameters as the typical values, all parameters would be pushed to these typical values since they are the best-fit ones for that subject. After selecting a sensitive parameter subset for each subject, we selected the most frequent subset among all subjects. We applied the LASSO method to the experimental data

Fig. 4 Results of applying the FIM method (Appendix C) with $n_{min} = 2$ to the experimental data. The optimal number of sensitive parameters per subset is n_θ^* . The number of subsets (each with n_θ^* parameters) is $N_{candidate}$.

Table 3 Parameter estimates of the complex model with 12 estimated parameters (Nominal) and the reduced models with five sensitive parameters selected by FIM and LASSO methods ($\hat{\theta}^{\text{FIM}}$ and $\hat{\theta}^{\text{LASSO}}$, respectively). Parameter 95% confidence interval width was estimated for the complex and reduced models. Goodness of fit was measured by VAF. All values are given as mean $\pm$ standard deviation across subjects.

Parameters	Parameter estimates			95% confidence interval width		
	$\hat{\Theta}$ (Nominal)	$\hat{\theta}^{\text{FIM}}$	$\hat{\theta}^{\text{LASSO}}$	CI(Θ)	CI(θ^{FIM})	CI(θ^{LASSO})
$K_{\text{vis}} (\frac{\text{Nm}}{\text{rad}})$	428 ± 125	$279^{\text{a,b}} \pm 27$	401 ± 163	546 ± 128	$62^{\text{c,d}} \pm 19$	$236^{\text{c}} \pm 122$
$K_{\text{vcr}} (\frac{\text{Nm}^2}{\text{rad}})$	4860 ± 2205	—	$8055^{\text{a}} \pm 1371$	7274 ± 806	—	$3623^{\text{c}} \pm 1418$
$K_{\text{ccr}} (\frac{\text{Nm}}{\text{rad}})$	87.9 ± 55.3	$16.2^{\text{a,b}} \pm 11.9$	$36.4^{\text{a}} \pm 30.5$	123.0 ± 49.8	$18.2^{\text{c,d}} \pm 6.0$	$68.7^{\text{c}} \pm 36.9$
$\tau(\text{s})$	0.263 ± 0.027	$0.242^{\text{b}} \pm 0.024$	$0.226^{\text{a}} \pm 0.027$	0.034 ± 0.010	$0.020^{\text{c,d}} \pm 0.0710$	0.033 ± 0.010
$\tau_{1\text{A}}(\text{s})$	0.099 ± 0.063	0.069 ± 0.013	—	0.134 ± 0.036	$0.030^{\text{c}} \pm 0.015$	—
$\tau_{\text{CNS1}}(\text{s})$	0.675 ± 0.280	$0.918^{\text{a}} \pm 0.077$	—	0.544 ± 0.145	$0.138^{\text{c}} \pm 0.069$	—
$\tau_{\text{c}}(\text{s})$	1.733 ± 1.118	—	2.200 ± 1.256	3.086 ± 0.495	—	$1.736^{\text{c}} \pm 0.534$
$\tau_{\text{CNS2}}(\text{s})$	27.2 ± 16.2	—	—	48.7 ± 2.4	—	—
$\tau_{\text{MS1}}(\text{s})$	0.433 ± 0.381	—	—	0.932 ± 0.075	—	—
$\tau_{\text{MS2}}(\text{s})$	0.439 ± 0.325	—	—	0.877 ± 0.086	—	—
$B (\frac{\text{Nm}^2}{\text{rad}})$	1.690 ± 1.213	—	—	4.659 ± 0.144	—	—
$K (\frac{\text{Nm}}{\text{rad}})$	2.170 ± 1.736	—	—	4.758 ± 0.114	—	—
VAF (%)	82.26 ± 7.67	82.65 ± 7.57	81.87 ± 7.56	—	—	—

^aThe parameter estimate in the reduced model (with five estimated parameters selected by FIM and LASSO methods) was significantly different ($p \leq 0.05$) from that of the complex model (with 12 estimated parameters – nominal values).

^bThe parameter estimate in the FIM-selected parameters was significantly different ($p \leq 0.05$) from that of the LASSO-selected parameters.

^cThe parameter 95% confidence interval in the reduced model (from FIM or LASSO methods) was significantly narrower ($p \leq 0.05$) than that of the complex model.

^dThe parameter 95% confidence interval from the FIM method was significantly narrower ($p \leq 0.05$) than that from the LASSO method.

where n_{θ}^* was set to 5 to obtain comparable results. The selected subset was $\theta^{\text{LASSO}} = \{K_{\text{vis}}, K_{\text{ver}}, K_{\text{ccr}}, \tau, \tau_c\}$. θ^{LASSO} shared three parameters (out of 5) with θ^{FIM} .

The 95% CI of the three common parameters $\{K_{\text{vis}}, K_{\text{ccr}}, \tau\}$ were reduced substantially more by FIM than LASSO. The K_{vis} CI was reduced by LASSO to 43% and by FIM to 11% of the original CI (Fig. 5(a)). The K_{ccr} CI was reduced by LASSO to 56% and by FIM to 15% of the original CI (Fig. 5(b)). The τ CI was reduced by LASSO to 96% and by FIM to 59% of the original CI (Fig. 5(c)). The CI of the selected sensitive parameters (θ^{FIM} and θ^{LASSO}) were significantly narrower than the original CI of all parameters Θ (Fig. 5 and Table 3), except the τ^{LASSO} versus τ^{ALL} comparison ($p = 0.69$). Moreover, the CI of the common parameters $\{K_{\text{vis}}, K_{\text{ccr}}, \tau\}$ in θ^{FIM} were significantly narrower than those CI in θ^{LASSO} (Table 3).

Goodness of fit measure VAF was determined by estimating the selected sensitive parameters (θ^{FIM} and θ^{LASSO}) while fixing the remaining parameters to the mean of the nominal values (Table 3 and Fig. 2). FIM VAF and LASSO VAF (VAF^{FIM} and $\text{VAF}^{\text{LASSO}}$, respectively) were almost equal, where $\text{VAF}^{\text{FIM}} = 83\% \pm 8\%$ and $\text{VAF}^{\text{LASSO}} = 82\% \pm 8\%$ across subjects for five selected sensitive parameters. When estimating all 12 model parameters, goodness of fit was $\text{VAF}^{\text{ALL}} = 82\% \pm 8\%$. Statistical analysis did not show significant differences between VAF^{FIM} or $\text{VAF}^{\text{LASSO}}$ versus VAF^{ALL} ($p = 0.23$ and 0.33 , respectively).

Very different computation times ($\mathcal{T}$) of FIM and LASSO methods were measured. All computation times were found using the MATLAB function “timeit” running on a single core of Intel Xeon E5-2699 v3. The average computation times to build FIM and solve the FIM method was $\mathcal{T}^{\text{FIM}} = 48.75\text{s}$, while the average LASSO computation time was $\mathcal{T}^{\text{LASSO}} = 7981.6\text{s}$.

Significant differences in parameter estimates occurred between the 12 estimated parameters for the complex model and each of the five estimated parameters selected by FIM and LASSO methods for reduced models. Three out of five parameters were significantly different from their nominal values: $\{K_{\text{vis}}, K_{\text{ccr}}, \tau_{\text{CNS1}}\}$ in θ^{FIM} , and $\{K_{\text{ver}}, K_{\text{ccr}}, \tau\}$ in θ^{LASSO} were different from their counterparts in Θ (Table 3). Moreover, the three parameters selected by both FIM and LASSO methods $\{K_{\text{vis}}, K_{\text{ccr}}, \tau\}$ had significantly different estimates.

4 Discussion

In this study, we compared our novel FIM method to the LASSO method for selecting sensitive model parameters based on four performance measures: (1) number of sensitive parameters that are in-common between the FIM and LASSO methods, (2) parameter CI, (3) VAF, and (4) computation time. First, the selected sensitive parameters, using FIM and LASSO methods, were quite similar. Three out of five selected parameters were identical $\{K_{\text{vis}}, K_{\text{ccr}}, \tau\}$. Thus, sensitive parameter selection is considered robust to the method used. Second, the 95% CI of the three common parameters $\{K_{\text{vis}}, K_{\text{ccr}}, \tau\}$ were reduced significantly by both FIM and LASSO methods versus the original CI, except the τ^{LASSO} case (Fig. 5). This indicates the success of both methods in selecting sensitive model parameters. Moreover, the CI were reduced significantly more by FIM than LASSO (Table 3), which demonstrates that our novel FIM method is superior to the LASSO method in reducing parameter 95% CI. Third, goodness of fit measure VAF for FIM or LASSO methods did not show significant differences from that of estimating all 12 model parameters. This indicates that the two methods maintained goodness of fit although they reduced the number of estimated parameters. Fourth, the average computation time of FIM method was less than 1% of that in LASSO method. Our novel FIM method is significantly faster than the LASSO method.

Significant differences in parameter estimates exist between the complex model with 12 estimated parameters (Θ) and the reduced models with five estimated parameters (θ^{FIM} and θ^{LASSO}). The reason might be that estimating all 12 parameters resulted in reaching a rather randomly found local minimum out of multiple local minima, due to poor identifiability, while estimating only sensitive parameters ensured the unique solution with good identifiability. Moreover, significant differences exist between estimates of the common parameters in θ^{FIM} and θ^{LASSO} . These differences are a result of selecting different sensitive parameters whose estimates might be biased. Simulating a hypothetical subject, whose physiological parameters are known, showed that there was a bias in the estimates of the selected sensitive parameters. Once a sensitive parameter selection method is chosen for a study (e.g., a comparison study of subjects' parameters before and after a therapy), the bias effect will be relatively

Fig. 5 Comparing the 95% confidence interval (CI) between estimating ALL parameters, FIM-selected parameters (θ^{FIM}), and LASSO-selected parameters (θ^{LASSO}). • is the mean point-estimate and the error bars are the mean CI (across subjects) for (a) K_{vis} , (b) K_{ccr} , and (c) τ that were selected as sensitive parameters by both FIM and LASSO methods.

constant throughout the study. Then, our approach can provide otherwise impossible comparison results especially for a parametric model with poor identifiability. Future research around parametric estimation of sensorimotor control should investigate these issues.

5 Conclusion

- Sensitive parameter selection was robust to the method used. Both FIM and LASSO methods selected quite similar sensitive parameters (three out of five selected parameters were identical).
- Both FIM and LASSO methods reduced the parameter 95% confidence intervals significantly up to 11% and 43% of the original CI, respectively. The FIM method reduced CI significantly more than LASSO.
- Both methods maintained the original goodness of fit. Across FIM and LASSO methods, goodness of fit measure was essentially equal (82%, Fig. 2).
- The FIM method is 164 times faster than the LASSO method.

This information will guide modelers on the pros and cons of each method and indicates that FIM more efficiently reaches an appropriate reduced model parameter set.

Acknowledgment

The contents of this work are solely the responsibility of the authors and do not necessarily represent the official views of NCCIH.

Funding Data

- National Center for Complementary and Integrative Health (NCCIH) at the National Institutes of Health (Grant No. U19AT006057).

Appendix A: Fisher Information Matrix (FIM)

Fisher information is a measure of the sensitivity of an observable random variable y relative to some parameter Θ . The experimental data value is more sensitive to the sensitive parameters. We used FIM to estimate sensitivity of each subset of the parameters as measured by estimation error variance.

Fisher information matrix was derived in Ref. [12] for the discrete-time system (given in state space)

$$\left. \begin{aligned} x(k+1) &= A(\Theta_0)x(k) + B(\Theta_0)u(k), \\ y(k) &= C(\Theta_0)x(k) + \omega(k), \forall k = 1, 2, \dots, N \end{aligned} \right\} \quad (A1)$$

where Θ_0 is the true parameters, y is the measurement output, $\omega \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_\omega)$ is a Gaussian measurement noise, and N is the number of observations. The FIM computed at Θ_0 is

$$\mathbb{I}(\Theta_0) = \sum_{k=1}^N \left(\frac{\partial e(k, \Theta)}{\partial \Theta} \bigg|_{\Theta=\Theta_0} \right)^T \Sigma_\omega^{-1} \left(\frac{\partial e(k, \Theta)}{\partial \Theta} \bigg|_{\Theta=\Theta_0} \right) \quad (A2)$$

where $\hat{y}(k, \Theta) = C(\Theta)x(k)$ is the model output, and $e = y - \hat{y}$ is the error. The derived FIM formula is [12]

$$\mathbb{I}(\Theta_0) = (\mathcal{H}\mathcal{U})^T (I_N \otimes \Sigma_\omega^{-1}) (\mathcal{H}\mathcal{U}), \quad \left. \begin{aligned} \text{where } \mathcal{U} &= I_{n_\Theta} \otimes U, \mathcal{H} = [H_1 \ H_2 \ \dots \ H_{n_\Theta}], H_i = \frac{\partial G}{\partial \Theta_i} \bigg|_{\Theta=\Theta_0} \\ U &= \begin{bmatrix} x_0 \\ u_0 \\ \vdots \\ u_{N-1} \end{bmatrix}, G = \begin{bmatrix} CA & CB & \dots & 0 \\ CA^2 & CAB & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^N & CA^{N-1}B & \dots & CB \end{bmatrix} \end{aligned} \right\} \quad (A3)$$

where $[u_0^T \ u_1^T \ \dots \ u_{N-1}^T]^T$ is the input sequence, x_0 is the initial state, I is the identity matrix, and $\otimes$ is the Kronecker tensor product operator.

Table 4 The algorithm of sensitive parameter selection using the LASSO method

Input: (1) Experimental Data and Dynamical Model
(2) Vector of Normalized Parameter Typical Values ($\Phi_0 \in \mathbb{R}^{n_\Theta}$), where the pre-normalization parameter vector is Θ
(3) Desired Optimal Number of Sensitive Parameters (n_θ^*)

Output: (1) A Subset of the Sensitive Parameters θ^{LASSO}

```

1. NumParams = 0
2.  $\lambda = 0.13$ 
3. while NumParams! =  $n_\theta^*$  do
4.   NumParams = 0
5.   S (selection matrix) = zeros ( $n_\theta^*, n_\Theta$ )
6.   Solve Eq. (C4)
7.   for  $i = 1: n_\Theta$  do
8.     if  $(\Phi(i) - \Phi_0(i)) > 0.005$  then
9.       NumParams = NumParams + 1
10.    end if
11.    if NumParams  $\leq n_\theta^*$  then
12.      S(NumParams, i) = 1
13.    end if
14.  end for
15.  if NumParams! =  $n_\theta^*$  then
16.     $\lambda = \lambda + \frac{\text{NumParams} - n_\theta^*}{i + n_\theta^*}$ 
17.  end if
18. end while
19.  $\theta^{\text{LASSO}} = S\Theta$ 

```

The estimation error converges to a normal distribution asymptotically under mild conditions [13,26,27], that is

$$(\hat{\Theta}_N - \Theta_0) \xrightarrow{d} \mathcal{N}(0, \mathbb{I}^{-1}) \quad (\text{A4})$$

The FIM inverse is a lower bound of the covariance matrix of the estimated parameters if $\hat{\Theta}_N$ is an unbiased estimator, that is [13,28]

$$\text{Cov}(\hat{\Theta}_N) \geq \mathbb{I}^{-1} \quad (\text{A5})$$

where $\mathbb{I}^{-1}$ is known as Cramér-Rao lower bound (CRLB), $\text{Cov}(\cdot)$ returns the covariance matrix and the matrix inequality $\geq$ means that $\text{Cov}(\hat{\Theta}_N) - \mathbb{I}^{-1}$ is positive semidefinite. This lower bound ($\mathbb{I}^{-1}$) presents the achievable performance limit by any unbiased estimator.

$\mathbb{I}(\Theta_0)$ of the unnormalized parameters may be misleading because of the different parameters' scales. Therefore, we used FIM of the normalized parameters given by

$$\mathbb{I}(\Phi_0) = \sum_{k=1}^N \left(\left. \frac{\partial e(k, \Phi)}{\partial \Phi} \right|_{\Phi=\Phi_0} \right)^T \Sigma_\omega^{-1} \left(\left. \frac{\partial e(k, \Phi)}{\partial \Phi} \right|_{\Phi=\Phi_0} \right) \quad (\text{A6})$$

Appendix B: Least Absolute Shrinkage and Selection Operator

The LASSO algorithm was developed as a method to simultaneously estimate parameters and reduce model complexity by shrinking insensitive parameters to zero, thereby removing them from the model [29]. The LASSO algorithm is an ordinary least squares estimator with the addition of an ℓ_1 norm penalization on the parameter vector [30]. The use of ℓ_1 norm results in solutions where some parameters are shrunk exactly to zero [29]. This results in selecting parameters that are the most sensitive in the model characterization, while simultaneously providing estimates of these sensitive parameters.

The LASSO minimization, in its most general form, is

$$\Theta = \arg \min_{\Theta} \|y - \hat{y}(\Theta)\|_2^2 + \lambda \|\Theta\|_1 \quad (\text{B1})$$

where λ is a tuning parameter to trade-off between goodness of fit and the number of selected sensitive parameters, and $\|\cdot\|_1$ is the ℓ_1 norm. This method can be applied to our problem by tuning λ to exploit LASSO's capability of model complexity reduction. However, it should be noted that LASSO is a biased estimator. This is

a result of the entire vector of parameters being compressed, resulting in compressed estimates of the parameters which remain in the model.

Appendix C: Fisher Information Matrix and Least Absolute Shrinkage and Selection Operator Methods For Sensitive Parameter Selection

The goal here is to find the pool of candidate subsets $\theta \subseteq \Theta$ of sensitive parameters. Let $\Theta \in \mathbb{R}^{n_\Theta}$ and $\theta \in \mathbb{R}^{n_\theta}$, then

$$\theta = \left\{ S\Theta \left| \begin{array}{l} S \in \mathbb{R}^{n_\theta \times n_\Theta} \text{ with zeros every where except for} \\ \text{a single 1 per row such that S has a full rank} \end{array} \right. \right\} \quad (\text{C1})$$

where S is a selection matrix. For example, to select the first and last parameters of $\Theta \in \mathbb{R}^{12}$, the selection matrix is

$$S = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{2 \times 12}$$

C.1 Fisher Information Matrix Method. To compute FIM submatrices $\mathbb{I}(\phi_0)$, where $\phi_0 = S\Phi_0$, we did not repeat the calculations in Eq. (A6). Instead, we computed $\mathbb{I}(\Phi_0)$ once, then obtained the submatrix

$$\mathbb{I}(\phi_0) = S\mathbb{I}(\Phi_0)S^T \quad (\text{C2})$$

The FIM-based selection problem must fulfill three requirements. First, it should select sensitive parameters. Therefore, we bounded the estimation error variance $\mathbb{I}^{-1}$ from above. Second, there should be no numerical issues with the matrix inversion. Therefore, we set a lower bound on the reciprocal condition number of $\mathbb{I}$ [31]. Third, to consider goodness of fit, we decided to maximize the number of selected sensitive parameters. Consequently, the FIM-based selection problem is

$$\left. \begin{array}{l} \theta^{\text{FIM}} = \arg \max_{\phi=S\Phi, \Phi_i=\frac{\Theta_i - \mu_i}{\sigma_i}} n_\theta, \\ \text{subject to : } n_\theta = \text{length}(\phi), \\ n_\theta \geq n_{\min}, \\ \text{rcond}(\mathbb{I}(\phi_0)) \geq \varepsilon, \\ d_{\max}(\mathbb{I}^{-1}(\phi_0)) \leq \sigma_{\text{allowable}}^2 \end{array} \right\} \quad (\text{C3})$$

where $\text{length}(\cdot)$ returns the length of a vector, $\text{rcond}(\cdot)$ returns the reciprocal condition number, ε is the machine epsilon, and $d_{\max}(\cdot)$ returns the maximum diagonal element of the input square matrix. $n_{\min}$ and $\sigma_{\text{allowable}}^2$ are user-specified constants which are the minimum number of sensitive parameters and maximum allowable estimation error variance of any normalized parameter, respectively. Given $\mathbb{I}(\Phi_0)$, $n_{\min}$, and $\sigma_{\text{allowable}}^2$, the optimization is solved by investigating all possible parameter subsets ($\phi \subseteq \Phi$) to check their feasibility (satisfy the constraints). Then, the subset(s) with the largest number of parameters is selected. The solution is one or more subsets of sensitive parameters θ^{FIM} . Let cardinality of θ^{FIM} (number of different subsets) be $N_{\text{candidate}}$.

There are two important aspects of our solution to the problem shown in Eq. (C3). First, the problem is formulated in a generic way to fit any application. Therefore, the user should provide the normalized FIM $\mathbb{I}(\Phi_0)$, $n_{\min}$ and $\sigma_{\text{allowable}}^2$. Second, the problem is solved by investigating all possible combinations of the parameters (combinatorial optimization). This method is feasible in our application since it has 12 parameters. If this is not the case, the reader is referred to Ref. [32] for its relaxation.

The choices of $\sigma_{\text{allowable}}^2$ and $n_{\min}$ need to be made carefully. Initially, we recommend setting $n_{\min}$ as small as possible and trying different values of $\sigma_{\text{allowable}}^2$ that are usually between 0 and 1, because the normalized parameters $\in [-1, 1]$. As $\sigma_{\text{allowable}}^2$ increases, the solution (selected parameter subsets) either (1) remains the same, (2) has more subsets, i.e., larger $N_{\text{candidate}}$, while the optimal number of parameters per subset n_{θ}^* remains the same, or (3) has a larger n_{θ}^* . If the optimization problem is infeasible, i.e., does not yield a solution, the user should decrease $n_{\min}$ or increase $\sigma_{\text{allowable}}^2$.

C.2 Least Absolute Shrinkage and Selection Operator

Method. The LASSO method is

$$\Phi = \arg \min_{\Phi} \|y - \hat{y}(\Phi)\|_2^2 + \lambda \|\Phi - \Phi_0\|_1 \quad (\text{C4})$$

where Φ_0 is a vector containing the typical parameter values, λ is a tuning parameter to trade-off between goodness of fit and the number of selected sensitive parameters, and $\|\cdot\|_1$ is the ℓ_1 norm. As the estimated parameters get pushed closer to their typical values, some parameters will become the typical values. The remaining parameters, whose values are not close enough to the typical values, are considered the sensitive parameters. This leads to a reduction in the total number of parameters requiring estimation. The result is a single subset of sensitive parameters.

The selected sensitive parameters are the ones that do not get pushed close enough to their typical values ($\Phi(i) - \Phi_0(i) > 0.005$) at a given λ . This process is repeated, while updating λ , until LASSO yields the number of sensitive parameters n_{θ}^* defined by the user beforehand. The algorithm of the LASSO method is summarized in Table 4.

Least absolute shrinkage and selection operator is typically applied to linear models (output is linear in parameters). In our case, the output is nonlinear in parameters, requiring a different method to solve the minimization problem. The Nelder–Mead simplex algorithm was used to solve Eq. (C4). The algorithm was implemented through the MATLAB function “fminsearchbnd,” which gives us the ability to impose bounds on the estimated parameters.

The estimated parameters from Eq. (C4) become compressed so we did not rely on these parameter estimates. Instead, we used the LASSO method solely for sensitive parameter selection. Once selected, the sensitive parameters were estimated using nonlinear least squares (uncompressed estimates). This method is known as relaxed LASSO [33].

References

- [1] Forbes, P. A., De Bruijn, E., Schouten, A. C., Van Der Helm, F. C. T., and Happee, R., 2013, “Dependency of Human Neck Reflex Responses on the Bandwidth of Pseudorandom Anterior-Posterior Torso Perturbations,” *Exp. Brain Res.*, **226**(1), pp. 1–14.
- [2] Peng, G. C. Y., Hain, T. C., and Peterson, B. W., 1996, “A Dynamical Model for Reflex Activated Head Movements in the Horizontal Plane,” *Biol. Cybern.*, **75**(4), pp. 309–319.
- [3] Peterson, B. W., Choi, H., Hain, T. C., Keshner, E. A., and Peng, G. C. Y., 2001, “Dynamic and Kinematic Strategies for Head Movement Control,” *Ann. N. Y. Acad. Sci.*, **942**(1), pp. 381–393.
- [4] Chen, K. J., Keshner, E. A., Peterson, B. W., and Hain, T. C., 2002, “Modeling Head Tracking of Visual Targets,” *J. Vestib. Res.*, **12**(1), pp. 25–33.
- [5] Goodworth, A. D., and Peterka, R. J., 2009, “Contribution of Sensorimotor Integration to Spinal Stabilization in Humans,” *J. Neurophysiol.*, **102**(1), pp. 496–512.
- [6] Moorhouse, K. M., and Granata, K. P., 2007, “Role of Reflex Dynamics in Spinal Stability: Intrinsic Muscle Stiffness Alone is Insufficient for Stability,” *J. Biomech.*, **40**(5), pp. 1058–1065.
- [7] van Drunen, P., Maaswinkel, E., der Helm, F. C. T., van Dieën, J. H., and Happee, R., 2013, “Identifying Intrinsic and Reflexive Contributions to Low-Back Stabilization,” *J. Biomech.*, **46**(8), pp. 1440–1446.
- [8] van der Helm, F. C. T., Schouten, A. C., de Vlugt, E., and Brouwn, G. G., 2002, “Identification of Intrinsic and Reflexive Components of Human Arm Dynamics During Postural Control,” *J. Neurosci. Methods*, **119**(1), pp. 1–14.
- [9] Schouten, A. C., De Vlugt, E., van Hiltten, J. J. B., and Van Der Helm, F. C. T., 2008, “Quantifying Proprioceptive Reflexes During Position Control of the Human Arm,” *IEEE Trans. Biomed. Eng.*, **55**(1), pp. 311–321.
- [10] Yua, T. F., and J. Wilson, A., 2014, “A Passive Movement Method for Parameter Estimation of a Musculo-Skeletal Arm Model Incorporating a Modified Hill Muscle Model,” *Comput. Methods Programs Biomed.*, **114**(3), pp. e46–e59.
- [11] Lin, D. C., and Nichols, T. R., 2003, “Parameter Estimation in a Crossbridge Muscle Model,” *ASME J. Biomech. Eng.*, **125**(1), pp. 132–140.
- [12] Priess, M. C., Choi, J., Radcliffe, C., Popovich, J. M., Cholewicki, J., and Reeves, N. P., 2015, “Time-Domain Optimal Experimental Design in Human Seated Postural Control Testing,” *ASME J. Dyn. Syst. Meas. Control*, **137**(5), pp. 545011–545017.
- [13] Ljung, L., 1999, *System Identification: Theory for the User*, PTR Prentice Hall, Upper Saddle River, NJ.
- [14] Grandjean, T. R. B., Chappell, M. J., Yates, J. W. T., and Evans, N. D., 2014, “Structural Identifiability Analyses of Candidate Models for In Vivo Pitavastatin Hepatic Uptake,” *Comput. Methods Programs Biomed.*, **114**(3), pp. e60–e69.
- [15] Ljung, L., 2013, “Some Classical and Some New Ideas for Identification of Linear Systems,” *J. Control. Autom. Electr. Syst.*, **24**(1–2), pp. 3–10.
- [16] Rojas, C. R., Welsh, J. S., Goodwin, G. C., and Feuer, A., 2007, “Robust Optimal Experiment Design for System Identification,” *Automatica*, **43**(6), pp. 993–1008.
- [17] Morris, E. D., Saidel, G. M., Chisolm, G. M., and Chisolm 3rd, G. M., 1991, “Optimal Design of Experiments to Estimate LDL Transport Parameters in Arterial Wall,” *Am. J. Physiol. Circ. Physiol.*, **261**(3), pp. H929–H949.
- [18] Zeinali-Davarani, S., Choi, J., and Baek, S., 2009, “On Parameter Estimation for Biaxial Mechanical Behavior of Arteries,” *J. Biomech.*, **42**(4), pp. 524–530.
- [19] Oishi, M. M. K., TalebiFard, P., and McKeown, M. J., 2011, “Assessing Manual Pursuit Tracking in Parkinson’s Disease Via Linear Dynamical Systems,” *Ann. Biomed. Eng.*, **39**(8), pp. 2263–2273.
- [20] Bozdogan, H., 2000, “Akaike’s Information Criterion and Recent Developments in Information Complexity,” *J. Math. Psychol.*, **44**(1), pp. 62–91.
- [21] Rasouli, M., Westwick, D. T., and Rosehart, W. D., 2012, “Reducing Induction Motor Identified Parameters Using a Nonlinear Lasso Method,” *Electr. Power Syst. Res.*, **88**, pp. 1–8.
- [22] Yun, Y., Kim, H.-C., Shin, S. Y., Lee, J., Deshpande, A. D., and Kim, C., 2014, “Statistical Method for Prediction of Gait Kinematics With Gaussian Process Regression,” *J. Biomech.*, **47**(1), pp. 186–192.
- [23] Ramadan, A., Choi, J., Radcliffe, C. J., Cholewicki, J., Reeves, N. P., and Popovich, J. M., 2017, “Robotic Solutions to Facilitate Studying Human Motor Control,” 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), Jeju, South Korea, June 28–July 1, pp. 174–178.
- [24] Popovich, J. M., Reeves, N. P., Priess, M. C., Cholewicki, J., Choi, J., and Radcliffe, C. J., 2015, “Quantitative Measures of Sagittal Plane Head-Neck Control: A Test–Retest Reliability Study,” *J. Biomech.*, **48**(3), pp. 549–554.
- [25] Garatti, S., and Bitmead, R. R., 2010, “On Resampling and Uncertainty Estimation in Linear System Identification,” *Automatica*, **46**(5), pp. 785–795.
- [26] Hjalmarsson, H., 2005, “From Experiment Design to Closed-Loop Control,” *Automatica*, **41**(3), pp. 393–438.
- [27] Aguero, J. C., and Goodwin, G. C., 2006, “On the Optimality of Open and Closed Loop Experiments in System Identification,” 45th IEEE Conference on Decision and Control (CDC), San Diego, CA, Dec. 13–15, pp. 163–168.
- [28] Emery, A. F., and Nenarokomov, A. V., 1998, “Optimal Experiment Design,” *Meas. Sci. Technol.*, **9**(6), pp. 864–876.
- [29] Tibshirani, R., 1996, “Regression Shrinkage and Selection Via the Lasso,” *J. R. Stat. Soc. Ser. B*, **58**(1), pp. 267–288.
- [30] Tibshirani, R., 2011, “Regression Shrinkage and Selection Via the Lasso: A Retrospective,” *J. R. Stat. Soc. Ser. B*, **73**(3), pp. 273–282.
- [31] Golub, G., and Loan, C. V., 2013, *Matrix Computations*, JHU Press, Baltimore, MD.
- [32] Joshi, S., and Boyd, S., 2009, “Sensor Selection Via Convex Optimization,” *IEEE Trans. Signal Process.*, **57**(2), pp. 451–462.
- [33] Meinshausen, N., 2007, “Relaxed Lasso,” *Comput. Stat. Data Anal.*, **52**(1), pp. 374–393.