

Identification of Multiple-Input Systems with Highly Coupled Inputs: Application to EMG Prediction from Multiple Intracortical Electrodes

David T. Westwick

dwestwic@ucalgary.ca

Department of Electrical and Computer Engineering, University of Calgary, Calgary, Alberta, T2N 1N4, Canada.

Eric A. Pohlmeier

e-pohlmeier@northwestern.edu

Department of Biomedical Engineering, Northwestern University, Evanston, IL 60208, U.S.A.

Sara A. Solla

solla@northwestern.edu

Department of Physiology, Northwestern Medical School, Chicago, IL, 60611, and Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208, U.S.A.

Lee E. Miller

lm@northwestern.edu

Department of Physiology, Northwestern Medical School, Chicago, IL 60611, U.S.A.

Eric J. Perreault

e-perreault@northwestern.edu

Department of Physical Medicine and Rehabilitation, Northwestern University Medical School, Chicago, IL 60611, U.S.A.

A robust identification algorithm has been developed for linear, time-invariant, multiple-input single-output systems, with an emphasis on how this algorithm can be used to estimate the dynamic relationship between a set of neural recordings and related physiological signals. The identification algorithm provides a decomposition of the system output such that each component is uniquely attributable to a specific input signal, and then reduces the complexity of the estimation problem by discarding those input signals that are deemed to be insignificant. Numerical difficulties due to limited input bandwidth and correlations among the inputs are addressed using a robust estimation technique based on singular value decomposition. The algorithm has been evaluated on both simulated and experimental data. The latter involved estimating the

relationship between up to 40 simultaneously recorded motor cortical signals and peripheral electromyograms (EMGs) from four upper limb muscles in a freely moving primate. The algorithm performed well in both cases: it provided reliable estimates of the system output and significantly reduced the number of inputs needed for output prediction. For example, although physiological recordings from up to 40 different neuronal signals were available, the input selection algorithm reduced this to 10 neuronal signals that made significant contributions to the recorded EMGs.

1 Introduction

Recent advances in microelectrode array technology have made it possible to record from multiple neurons simultaneously (Maynard, Nordhausen, & Normann, 1997; Williams, Rennaker, & Kipke, 1999; Nicoletis et al., 2003). This capability may enhance our understanding of communication within the central nervous system (CNS) and may allow the development of brain-machine interfaces (BMIs) that provide enhanced communication and control for individuals with significant neurological disorders (Donoghue, 2002; Mussa-Ivaldi & Miller, 2003). However, the potential of these recording devices is limited by the current methods available for processing multi channel recordings. The purpose of this study is to develop robust and efficient algorithms for determining the dynamic relationship between a set of neural recordings and a continuous time signal related to those recordings.

Much of the research incorporating multielectrode arrays has focused on using intra cortical recordings to predict kinematic and dynamic features of hand motion in freely moving primates and on using these predictions as a basis for developing cortical BMIs. To date, a number of linear and nonlinear algorithms have been used to generate the map between cortical activity and specific movement variables (Serruya, Hatsopoulos, Paninski, Fellows, & Donoghue, 2002; Taylor, Tillery, & Schwartz, 2002; Carmena et al., 2003). In general, linear models have been found to perform nearly as well as nonlinear models for the prediction of continuous movement variables (Wessberg et al., 2000; Gao, Black, Bienenstock, Wu, & Donoghue, 2003). However, nonlinear models can have advantages for predicting the rest periods between movements or the peak velocities during the fastest movements in a continuous movement sequence (Kim et al., 2003). As might be expected, performance of either type of model generally improves with increasing numbers of neurons. Prediction accuracy for various movement-related signals has been shown to increase with increasing numbers of neurons for as many as 250 simultaneously recorded cortical neurons (Carmena et al., 2003). However, (Sanchez et al., 2004) recently demonstrated that prediction accuracy can be improved further by an appropriate selection of inputs.

Two main sources of error can arise when a large number of neurons are used as inputs to the system identification process. The first is that

correlations among neurons can result in a numerically ill-conditioned estimation problem. The second is that using too many inputs can lead to accurate fits of the data employed in the estimation process but poor generalization to new data sets. Although additional neural recordings can provide novel information, not all of this information may be relevant to the task or process under study. Therefore, techniques that reduce the dimensionality of the input signals could reduce the computational complexity of the system identification problem and possibly increase prediction accuracy. One approach to this problem has been to use principal component analysis (PCA) to generate a set of orthogonal inputs that span the space defined by the original data (Chapin, Moxon, Markowitz, & Nicoletis, 1999; Isaacs, Weber, & Schwartz, 2000; Wu et al., 2003). Such techniques can reduce the dimensionality of the input signal space when there are correlations between the input signals. This reduction can enhance the robustness of the identification process, but it does not reduce the number of neural signals that need to be recorded, since each principal component is a linear combination of all available input signals. An alternative approach is to select the "most relevant" set of inputs (Sanchez et al., 2004). Once this selection is complete, spike sorting and subsequent signal processing can be restricted to the retained inputs.

Input signals with restricted bandwidths also can lead to a numerically unstable identification problem and poor generalization. This problem manifests itself as the need to invert an ill-conditioned input correlation matrix, a problem that can be alleviated by using standard singular value decomposition (SVD) techniques to check for numerical instabilities (Paninski, Fellows, Hatsopoulos, & Donoghue, 2004) during the inversion process. When such instabilities exist, robust estimates of the impulse response functions (IRFs) between the inputs and the output still may be obtained by using an SVD-based matrix pseudo-inverse (Westwick & Kearney, 1997b).

The goal of this study is to develop robust tools for processing information obtained from large numbers of neural recordings and for determining the linear relationship between these recordings and a related physiological signal of interest. Specifically, we have developed an algorithm for selecting an optimal set of input signals, based on their unique contributions to the system output, and for developing robust predictors of the output from this subset of inputs. The performance of these novel methods is demonstrated on both simulated and experimental data; the latter consist of intracortical recordings from the primary motor cortex of a freely moving primate together with EMG data taken from several arm muscles.

2 Analytical Methods

Consider a multiple-input single-output (MISO) system, represented by a bank of N linear finite impulse response (FIR) filters with memory length

M . Let $x_k(t)$, for $k = 1, 2, \dots, N$, be measurements of the N input signals at time t , and let $z(t)$ be the measured output. Then,

$$z(t) = \sum_{k=1}^N \sum_{\tau=0}^{M-1} h_k(\tau) x_k(t - \tau) + w(t), \quad (2.1)$$

where $w(t)$ accounts for both noise in the output measurements and the effects of any additional, unmeasured inputs to the system. It is assumed to be zero mean and uncorrelated with all the measured inputs $x_k(t)$, $k = 1, \dots, N$.

The objective is to estimate the NM filter weights, $h_k(\tau)$ for $\tau = 0, \dots, M - 1$ and $k = 1, \dots, N$, from input-output measurements $x_k(t)$, $z(t)$, for $t = 1, \dots, T$. Given sufficient data, ideally $T \gg NM$, this can be accomplished by rewriting equation 2.1 as a matrix equation,

$$\bar{z} = \mathbf{X}\bar{h} + \bar{w}, \quad (2.2)$$

where \bar{z} and \bar{w} are T element vectors containing $z(t)$ and $w(t)$, respectively. The IRFs $h_k(\tau)$ are placed, in order, in the NM element vector \bar{h} ,

$$\bar{h} = [h_1(0) h_1(1), \dots, h_1(M-1) h_2(0) h_2(1), \dots, h_N(M-1)]^T. \quad (2.3)$$

Thus, \mathbf{X} is the block structured matrix,

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N], \quad (2.4)$$

where the \mathbf{X}_k are $T \times M$ matrices,

$$\mathbf{X}_k = \begin{bmatrix} x_k(1) & 0 & \dots & 0 \\ x_k(2) & x_k(1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x_k(T) & x_k(T-1) & \dots & x_k(T-M+1) \end{bmatrix}.$$

Since the noise is uncorrelated with the inputs, the minimum mean squared error estimate of \bar{h} can be obtained using the normal equation (Golub & Van Loan, 1989),

$$\hat{\bar{h}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \bar{z}. \quad (2.5)$$

One disadvantage of the direct solution, equation 2.5, of the normal equation is that the matrices can become unacceptably large. For example, in

the neural processing experiment described in section 4.2, the system had 40 inputs, each filter was represented using a 52 tap FIR filter, and the identification was performed using up to 18,000 data points. Thus, direct application of equation 2.5 would require multiplying an 18,000 by 2080 element matrix with its transpose, and then computing the inverse of the resulting 2,080 by 2080 matrix.

2.1 Auto- and Cross-Correlations. Perreault, Kirsch, and, Acosta (1999) developed an efficient solution to the MISO system identification problem based on auto- and cross-correlation functions instead of direct use of the data. They have shown that the input-output relationship can be rewritten in terms of auto- and cross-correlation matrices,

$$\begin{bmatrix} \vec{\phi}_{x_1z} \\ \vdots \\ \vec{\phi}_{x_Nz} \end{bmatrix} = \begin{bmatrix} \Phi_{x_1x_1} & \dots & \Phi_{x_1x_N} \\ \vdots & \ddots & \vdots \\ \Phi_{x_Nx_1} & \dots & \Phi_{x_Nx_N} \end{bmatrix} \begin{bmatrix} \vec{h}_1 \\ \vdots \\ \vec{h}_N \end{bmatrix}, \tag{2.6}$$

where \vec{h}_k is a vector whose elements are the samples of the IRF $h_k(\tau)$, $\vec{\phi}_{x_kz}$ is an M element vector containing the cross-correlation $\phi_{x_kz}(\tau)$, and $\Phi_{x_kx_\ell}$ is an $M \times M$ Toeplitz structured matrix whose elements are the correlations $\Phi_{x_kx_\ell}(i, j) = \phi_{x_kx_\ell}(i - j)$. In compact notation, equation 2.6 is written as

$$\vec{\Phi}_{Xz} = \Phi_{XX}\vec{h}. \tag{2.7}$$

Perreault et al. (1999) estimated the IRFs of the linear filters by solving equation 2.7 exactly through an inversion of the matrix Φ_{XX} ,

$$\hat{h} = \Phi_{XX}^{-1}\vec{\Phi}_{Xz}, \tag{2.8}$$

and noticed that the input correlation matrix Φ_{XX} could become ill conditioned if either the input signals are strongly coupled with one another or they are severely band-limited.

The equivalence of the two algorithms becomes evident when noting that

$$\frac{1}{T}\mathbf{X}^T\mathbf{X} = \Phi_{XX} + O\left(\frac{M}{T}\right), \tag{2.9}$$

$$\frac{1}{T}\mathbf{X}^T\vec{z} = \vec{\Phi}_{Xz}, \tag{2.10}$$

where $O(M/T)$ indicates an error with magnitude of the same order as M/T . Thus, if $T \gg M$, the solutions \hat{h} obtained from equations 2.5 and 2.8

will be virtually identical. Note that the error term of order M/T appears in the quadratic factor, equation 2.9 but not in the linear factor, equation 2.10. Furthermore, the corrections needed to make equation 2.9 exact can be implemented using the method suggested by Korenberg (1988), although the effect of these corrections is negligible unless the model's memory length M is significant compared to the data length T .

2.2 Input Contributions to the Output. The neural input data may contain signals that are either unrelated to the output variable of interest or are highly correlated with one another. Both scenarios lead to increases in the variability of the estimated model and decreases in its ability to generalize. To address these problems, an algorithm was developed that locates and eliminates redundant or irrelevant inputs. The algorithm is a variation on the orthogonal least-squares technique (Chen, Cowan & Grant, 1991), but it is based on a backward elimination approach rather than on forward regression (Miller, 1990). Thus, at each stage, this iterative algorithm computes the unique contribution that each input makes to the output and then eliminates the input that makes the smallest such contribution.

To find the component in the output that can be attributed only to the input $x_k(t)$, construct the matrices

$$\mathbf{M}_1 = [\mathbf{X}_1, \dots, \mathbf{X}_{k-1}, \mathbf{X}_{k+1}, \dots, \mathbf{X}_N],$$

$$\mathbf{M}_2 = \mathbf{X}_k,$$

orthogonalize \mathbf{M}_2 , which contains delayed copies of the input $x_k(t)$, against the remaining inputs, stored in \mathbf{M}_1 , and then project the output $z(t)$ onto these orthogonal columns using the the following QR factorization (Golub & Van Loan, 1989)

$$\begin{aligned} [\mathbf{M}_1 \quad \mathbf{M}_2 \quad \vec{z}] &= \mathbf{Q}\mathbf{R} \\ &= [\mathbf{Q}_1 \quad \mathbf{Q}_2 \quad \vec{q}_z] \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \vec{r}_{1z} \\ 0 & \mathbf{R}_{22} & \vec{r}_{2z} \\ 0 & 0 & r_{3z} \end{bmatrix}, \end{aligned} \quad (2.11)$$

where $\mathbf{Q}^T\mathbf{Q}$ is the $NM \times NM$ identity matrix and \mathbf{R} is upper triangular by construction. These matrices are partitioned such that \mathbf{Q}_1 and \mathbf{Q}_2 have the same dimensions as \mathbf{M}_1 and \mathbf{M}_2 , respectively. The dimensions of the blocks in \mathbf{R} can be inferred from

$$\begin{aligned} \mathbf{M}_1 &= \mathbf{Q}_1\mathbf{R}_{11} \\ \mathbf{M}_2 &= \mathbf{Q}_1\mathbf{R}_{12} + \mathbf{Q}_2\mathbf{R}_{22} \\ \vec{z} &= \mathbf{Q}_1\vec{r}_{1z} + \mathbf{Q}_2\vec{r}_{2z} + r_{3z}\vec{q}_z. \end{aligned} \quad (2.12)$$

Since the columns of \mathbf{Q} are orthogonal, the three terms on the right-hand side of equation 2.12 are orthogonal to each other. Thus, $\mathbf{Q}_2\vec{r}_{2z}$, which is orthogonal to \mathbf{Q}_1 and hence \mathbf{M}_1 , is the component in the output that can be attributed only to the input $x_k(t)$. The mean squared value of this unique contribution, $\hat{y}_k(t)$, is given by:

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \hat{y}_k^2(t) &= \frac{1}{T} (\mathbf{Q}_2\vec{r}_{2z})^T (\mathbf{Q}_2\vec{r}_{2z}) \\ &= \frac{\vec{r}_{2z}^T \vec{r}_{2z}}{T}. \end{aligned} \quad (2.13)$$

The procedure described above can be used to identify the input that makes the smallest unique contribution to the output. The least significant input may then be removed from the pool of inputs and the process repeated. Note that if there are correlations between the inputs, the significance of the remaining inputs will change as a result of the deletion. Hence, it is necessary to repeat this process $(N - 1)$ times to determine an optimal set of inputs to use in the identification process.

Since the computational cost associated with the QR factorization in equation 2.11 is approximately $4T(MN)^2$ flops (Golub & Van Loan, 1989) and since this factorization would have to be repeated once for each input, this scheme is clearly not practical. However, direct computation of the QR factorization is not necessary. To simplify the notation, consider estimating the contribution due to the last input $x_N(t)$, so that the QR factorization in equation 2.11 involves the matrix

$$[\mathbf{M}_1 \quad \mathbf{M}_2 \quad \vec{z}] = [\mathbf{X} \quad \vec{z}].$$

Squaring the right-hand side of equation 2.11 yields

$$(\mathbf{QR})^T \mathbf{QR} = \mathbf{R}^T \mathbf{R}, \quad (2.14)$$

while squaring the left-hand side gives

$$[\mathbf{X} \quad \vec{z}]^T [\mathbf{X} \quad \vec{z}] = \begin{bmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \vec{z} \\ \vec{z}^T \mathbf{X} & \vec{z}^T \vec{z} \end{bmatrix} \quad (2.15)$$

$$= T \begin{bmatrix} \Phi_{XX} & \vec{\Phi}_{Xz} \\ \vec{\Phi}_{Xz}^T & \sigma_z^2 \end{bmatrix}, \quad (2.16)$$

where the last equality follows from equations 2.9 and 2.10. Thus, the matrix \mathbf{R} , and hence the mean-squared value of $\hat{y}_N(t)$, can be obtained by computing the Cholesky factorization (Golub & Van Loan, 1989) of the matrix in

equation 2.16, which is constructed from the auto- and cross-correlation matrices. Note that the computational cost of the Cholesky factorization is approximately $(1/3)(NM)^3$ flops, independent of T . Clearly, the contribution due to any one of the inputs can be obtained by rearranging the blocks in Φ_{XX} and Φ_{Xz} so that the input of interest appears in the bottom and right-most block of rows and columns in Φ_{XX} and in the bottom-most block of rows in Φ_{Xz} .

2.3 Singular Value Decomposition. Although the input selection algorithm will reduce correlations between inputs, the linear regression may still be poorly conditioned. For example, input properties such as limited bandwidth, which will produce an ill-conditioned autocorrelation matrix, are not altered by the selection algorithm. This ill-conditioned regression problem can be solved robustly using the singular value decomposition (SVD) (Golub & Van Loan, 1989) of the regression matrix \mathbf{X} ,

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \quad (2.17)$$

where $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}$, and $\mathbf{S} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_{NM})$, with $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{NM} \geq 0$. Consider the estimate:

$$\hat{\mathbf{h}} = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^T\bar{\mathbf{z}} \quad (2.18)$$

$$= \bar{\mathbf{h}} + \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^T\bar{\mathbf{w}} \quad (2.19)$$

$$= \mathbf{V}(\mathbf{V}^T\bar{\mathbf{h}} + \mathbf{S}^{-1}\mathbf{U}^T\bar{\mathbf{w}}). \quad (2.20)$$

Aside from finite-precision errors, the solutions in equations 2.5 and 2.18 are identical. However, rewriting equation 2.18 as 2.19 provides insight into the effect that measurement noise will have on the final estimate. Thus, let

$$\bar{\boldsymbol{\eta}} = \mathbf{U}^T\bar{\mathbf{w}} \quad (2.21)$$

$$\bar{\boldsymbol{\zeta}} = \mathbf{V}^T\bar{\mathbf{h}}, \quad (2.22)$$

and let η_k and ζ_k be the k th elements of the vectors $\bar{\boldsymbol{\eta}}$ and $\bar{\boldsymbol{\zeta}}$, respectively. The estimate $\hat{\mathbf{h}}$ can then be written as

$$\hat{\mathbf{h}} = \sum_{k=1}^{NM} \left(\zeta_k + \frac{\eta_k}{\sigma_k} \right) \bar{\mathbf{v}}_k, \quad (2.23)$$

where $\bar{\mathbf{v}}_k$ is the k th column of the matrix \mathbf{V} . The decomposition in equation 2.23 expands the vector of IRF estimates as a linear combination of an orthogonal basis formed by the right singular vectors. Each expansion

coefficient consists of two terms: ζ_k , the projection of the true system onto the k th right singular vector, and η_k/σ_k , the projection of the measurement noise onto the k th left singular vector. Note, however, that the k th noise term is scaled by $1/\sigma_k$. Thus, small singular values can be expected to produce relatively large noise terms, and hence large errors in \hat{h} . The goal is to retain only those terms in equation 2.23 that are dominated by the signal component and to discard the rest (Westwick & Kearney, 1997b).

One approach to the selection of significant terms is to reorder the singular values according to their contributions to the output. The model can then be built up term by term, including the most significant remaining term at each step. Once an acceptable level of model accuracy has been reached, the expansion can be halted.

The model output is given by

$$\begin{aligned} \hat{y} &= X\hat{h} \\ &= USV^T\hat{h}. \end{aligned} \tag{2.24}$$

Define the coefficient vector,

$$\vec{\gamma} = SV^T\hat{h}, \tag{2.25}$$

which contains the projection of the IRF estimate onto the right singular vectors, scaled by their associated singular values. The mean squared value of the model output is then

$$\frac{1}{T}\hat{y}^T\hat{y} = \frac{1}{T}\vec{\gamma}^T\vec{\gamma}. \tag{2.26}$$

Thus, γ_k^2 , the square of the k th element of the vector $\vec{\gamma}$, represents the contribution made by the k th term in equation 2.23 to the mean square of the model output (Westwick & Kearney, 1997a). To sort the terms in order of decreasing significance, we need only to sort them in decreasing absolute value of the γ_k .

Finally, we note that the SVD may be used in conjunction with the efficient correlation-based technique proposed by Perreault et al. (1999). Calculate the SVD of the input correlation matrix,

$$\Phi_{XX} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T, \tag{2.27}$$

where $\mathbf{\Lambda} = \frac{1}{T}\mathbf{S}^2$. The initial estimate of the IRFs, based on equation 2.8, is then

$$\hat{h} = \sum_{k=1}^{NM} \frac{1}{\lambda_k} \vec{v}_k \vec{v}_k^T \vec{\Phi}_{Xz}. \tag{2.28}$$

The terms in equation 2.28 can be sorted in decreasing order of contribution to the output; the mean squared value of the output contribution of each of these terms can be calculated using

$$\gamma_k^2 = \lambda_k \left(\vec{v}_k^T \hat{h} \right)^2. \quad (2.29)$$

The mean squared output is then plotted versus the number of singular vectors retained in the model. The point where the plot starts to level off determines how many singular vectors to include in the final model.

3 Experimental Methods

The algorithms described above have been evaluated using data from both computer simulations and physiological recordings. The rationale behind the use of artificial data is to test our algorithms on a system with known and modifiable properties; the application to physiological recordings then demonstrates the performance of these algorithms under more realistic conditions, where the system under study may not be well characterized a priori.

3.1 Generation of Artificial Data. The input selection and system identification algorithms were evaluated on a simulated linear MISO system with highly correlated and band-limited inputs. Both of these characteristics significantly complicate the identification process and are likely to be encountered when recording a large number of physiological signals. A schematic representation of the simulation process used to generate the artificial input-output data is shown in Figure 1A.

The inputs were generated using K normally distributed independent white noise sources. Each of these sources was band-limited using a digital Butterworth filter with a randomly generated order, ranging from first to fourth, and a randomly generated cut-off frequency, ranging from 10% to 90% of the Nyquist rate. Unique filters were used for each input. The resulting set of K -independent, band-limited sources was multiplied by a randomly generated $K \times N$ mixing matrix, resulting in N correlated signals. In all cases N was greater than K , to mimic the recording of a large number of physiological signals driven by a small number of independent sources. Our goal was to evaluate the system identification procedure both with and without the optimal selection algorithm. Since this identification algorithm will not work if the inputs are fully statistically dependent, N independent gaussian white noise sources were added to the N dependent signals, to ensure some degree of independence. A 10 dB signal to noise ratio was used to emphasize the coupling among the N generated inputs over their small degree of independence.

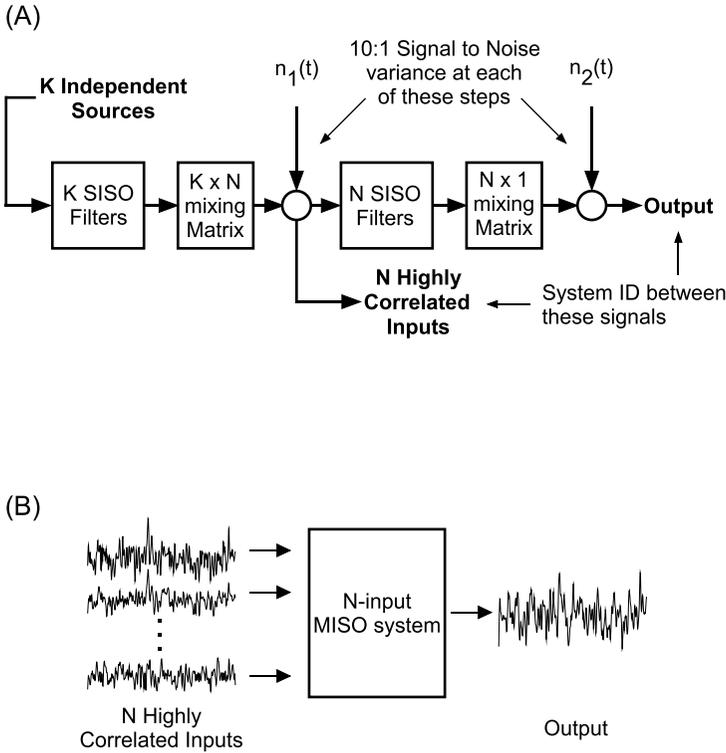


Figure 1: Schematic representation of the computer simulations. (A) Diagram of the process used to generate artificial data. The K independent sources are gaussian white noise. The SISO filters are digital Butterworth filters with randomly generated orders and cut-off frequencies. Specific details are provided in the text. (B) Typical input and output signals generated by this process; 250 sample points are shown.

The system output was generated from these inputs using a similar process. Each one of the N inputs was filtered by a unique digital Butterworth filter with a randomly generated order, ranging from first to fifth, and a randomly generated cutoff frequency, ranging from 10% to 80% of the Nyquist rate. These parameters were chosen so that on average, the bandwidth of the system inputs was greater than that of the system to be identified. The resulting signals were combined using a randomly generated $N \times 1$ mixing matrix to produce a single output. Measurement noise was simulated by adding normally distributed white noise with 10 dB signal-to-noise ratio. Figure 1B shows typical input and output signals generated by this process.

Monte Carlo simulations were used to evaluate algorithm performance. Each set of simulations used a fixed number of independent sources (K)

and corresponding inputs (N). In contrast, different noise sources, filter parameters, and mixing matrices were used for each trial in the set. Specific values for these stochastic model parameters were selected as described above. Unless specified, each Monte Carlo simulation set consisted of 100 trials, to obtain robust estimates of the mean and standard deviation for each quantity of interest.

3.2 Physiological Recordings. The algorithms also were evaluated using a set of physiological recordings obtained from a single behaving primate (*Macacca mulatta*) during execution of a button-pressing task. A total of 29 3-minute data files were collected during eight separate recording sessions. The recording sessions spanned a period of three months; it is assumed that the neural signals recorded from each electrode varied over the course of this period. Each trial began with the left hand on a touch pad at waist level. Following a 0.3 second touch pad hold time, one of four buttons in front of the subject was illuminated, instructing the subject to reach toward and press this randomly chosen illuminated button. The buttons were arranged on the top and bottom surfaces of a plastic box (see Figure 2A), thus requiring the subject to approach the button with the forearm either pronated or supinated, respectively. After a brief hold (~ 200 ms), a tone indicated success; the subject was given a juice reward and could return its hand to the touch pad to initiate the next trial after a random intertrial interval. Figure 2B shows data, including associated neuronal discharge signals, electromyograms (EMGs) from four muscles, and a binary trace indicating the button press times for a series of five such trials. There were an average of 60 ± 8 button presses in each 3 minute trial, equally distributed across the four targets.

The neuronal discharge signals were recorded from an array of 100 electrodes (Cyberkinetics, Inc.) The array was chronically implanted under the dura in the arm area of the primary motor cortex. Leads from the array were routed to a connector implanted on the skull. Signals from the best 32 of these electrodes were sent to a DSP-based multineuron acquisition system (Plexon, Inc, Dallas, TX) for later analysis with the Plexon Off-line Sorter software. For the data presented here, the sorting algorithm was able to distinguish between 35 to 40 independent neural signals from the 32 electrode recordings. It was possible to classify approximately 15% of these signals as single neurons, based on stringent shape and minimum interspike interval criteria. The remaining signals were those for which action potentials probably were due to more than one neuron, but from which background noise and occasional artifacts were removed. The spike occurrences for these discriminated signals were then converted into a rate code, sampled at 100 Hz, for subsequent processing.

EMG signals were recorded from surface electrodes placed above the anterior deltoid (AD), biceps (BI), triceps (TRI), and combined wrist and digit flexors (FF). The signals were sampled at 2000 Hz and subsequently

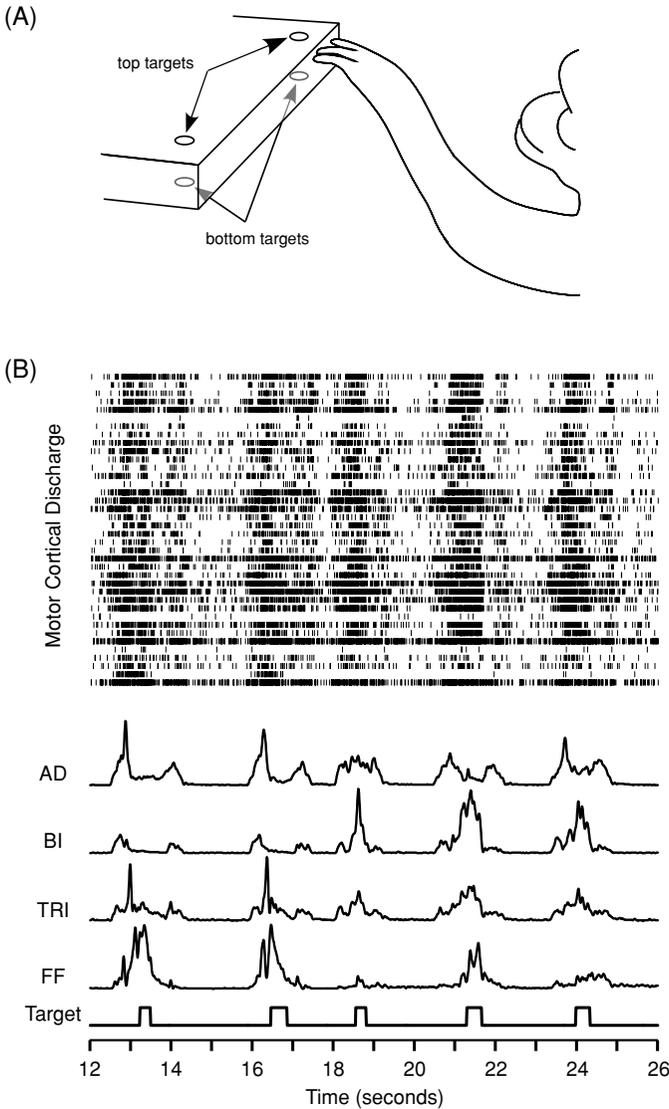


Figure 2: Experimental setup for physiological recordings. (A) In this reaching task, the monkey is required to press one of four lighted buttons located on the top and bottom surfaces of the target platform. (B) Typical data recorded during this task. The top traces correspond to the neuronal firing patterns recorded from the intracortical microelectrode array. The bottom traces show the simultaneously recorded electromyograms from four of the arm muscles involved in the task: anterior deltoid (AD), biceps (BI), triceps (TRI), and combined wrist and digit flexors (FF). The target trace identifies periods of button pressing.

rectified, low-pass-filtered (10 Hz), and resampled at 100 Hz. All animal-related procedures were approved by the institutional Animal Care and Use Committee at Northwestern University.

3.3 Statistical Analysis. To quantify model accuracy, we use r^2 , the square of the correlation coefficient between the system and model outputs. Typically, experimental data are divided into two sets: estimation data, and validation data. The model is identified from the estimation data, and tested on the validation data. If the model is validated using the estimation data, the value of r^2 can be biased upward if the model fits some of the measurement noise (overfitting). Thus, the difference between the values of r^2 obtained from the estimation and validation data can be used to assess the degree of overfitting.

The accuracy of the proposed input selection algorithm was compared to that of PCA and of two of the methods recently proposed by Sanchez et al. (2004). All of these alternative algorithms reduce the dimensionality of the input signal space. PCA was chosen for comparison, since it is a commonly used approach for improving the numerical stability of the identification process. PCA differs from the Sanchez algorithms in that it uses all inputs in the prediction process. In contrast, the methods proposed by Sanchez are designed to remove unnecessary neural signals. The first method (fSISO) ranks each input according to the amount of variance it can predict in the system output. Prediction is performed using a SISO linear filter and ignoring all other available inputs; correlations among inputs are not considered. The second method (fMISO) ranks the inputs according to the sum of the estimated linear filter magnitudes; the estimated filters are obtained using a nonparametric linear MISO identification. This was the most effective of the three algorithms evaluated by Sanchez et al. (2004). As with the optimal selection algorithm, this ranking process must be repeated ($N - 1$) times for a system with N inputs. Each iteration involves eliminating the least significant input.

4 Results

4.1 System Identification from Artificial Data. In the case of artificial data, the optimal selection algorithm greatly reduced the number of inputs needed for an accurate prediction of the system output. The performance of our input selection algorithm was compared to that obtained if a subset of the input signals was randomly chosen for use in the system identification process. The value of r^2 calculated from the cross-validation data was used to compare the performance of these two selection processes. Figure 3 shows the average results from a set of 100 Monte Carlo simulations using 10 independent sources ($K = 10$) to produce 20 correlated inputs ($N = 20$). For these simulation parameters, more than 90% of the maximally obtainable output variance could be predicted using the three most significant inputs.

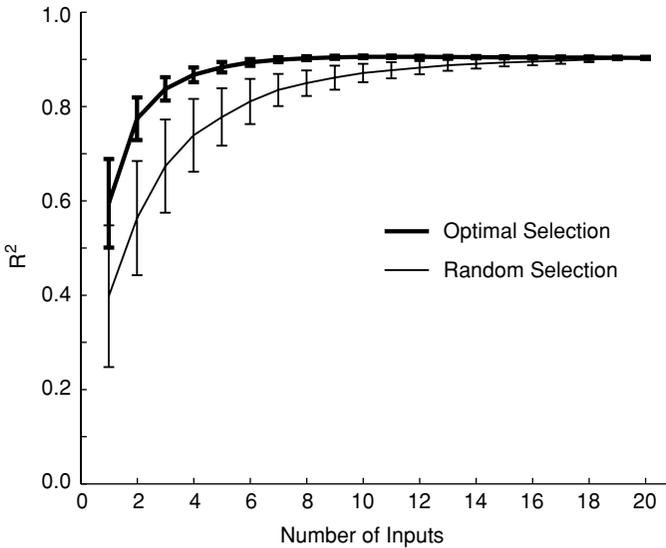


Figure 3: Model accuracy as a function of the number of inputs used in the identification process. Thin traces correspond to the value of r^2 for randomly selected inputs, and thick traces correspond to that for optimally selected inputs. Error bars indicate the standard deviation based on the results of 100 simulated trials. Simulation parameters: $K = 10$, $N = 20$, $T = 2000$ data points for each of the simulated data vectors. The estimated linear filters had lengths of $M = 32$ points.

In contrast, more than twice as many inputs were required to reach the same level of fitting accuracy when a random selection was used. Similar results were obtained for a wide range of simulation parameters, as examined by varying the number K of independent sources from 7 to 15 and the number N of input signals generated by these sources from 7 to 40.

Although it reduced the fitting accuracy, the robust identification algorithm improved output predictions for cross-validation data. The prediction accuracy of the estimated linear system was evaluated for both the fitted data and for cross-validation data not used in the fitting process. These results, as functions of the number of singular values used in the pseudoinverse, are shown in Figure 4. Simulation parameters were identical to those used for Figure 3, although half of the resulting data set was used for identification and the remaining half for cross-validation. As expected, r^2 for the fitted data increases monotonically with the number of singular values used in the identification process. However, there is a clear peak in the r^2 value for the cross-validation data, indicating the advantage of restricting the number of singular values when computing the pseudo-inverse for this

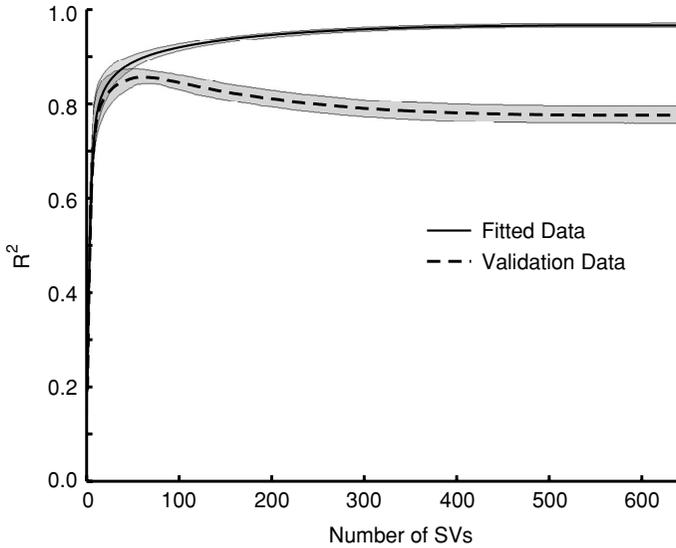


Figure 4: Prediction accuracy for the fitted (solid black trace) and cross-validation (dashed black trace) data as a function of the number of singular values used in the pseudo-inverse. Results are the average and standard deviation (gray bands) from 100 simulated data sets. Simulation parameters: $K = 10$, $N = 20$, 2000 data points for each of the simulated data vectors (1000 points were used for the estimation and 1000 for the cross-validation). The estimated linear filters had a length of $M = 32$ points.

data set. Such a result is a signature of overfitting, and the simulation parameters were chosen to illustrate this point. For the same set of simulation parameters, the discrepancy between the curves for the fitting and validation data diminishes as the number of data points used in the fitting process is increased.

The combined use of the optimal input selection algorithm and the robust MISO identification yields accurate output predictions using only a small subset of the measured inputs. Figure 5 shows typical model predictions for the same system used to generate the data for Figures 3 and 4. Only 3 of the 20 available inputs were used to predict the output response. Figure 5A shows the results when optimal input selection was combined with the robust identification algorithm. Based on the results shown in Figure 4, the number of singular values needed to predict 90% of the output variance was used to compute the pseudo-inverse. Figure 5B shows typical results using three randomly chosen inputs and all singular values in the identification process. For this data set, r^2 increased by 0.1 when the optimal input selection algorithm was used in conjunction with the robust MISO

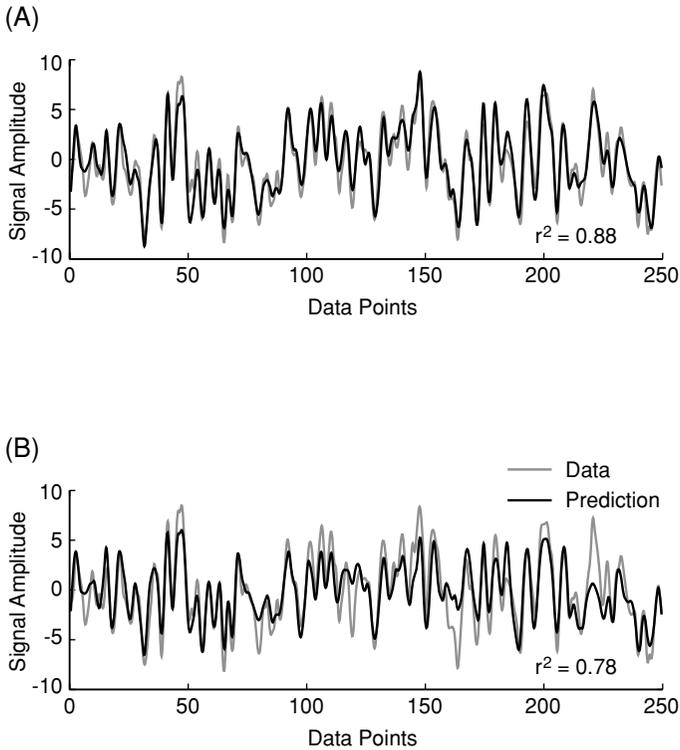


Figure 5: Actual (gray traces) and predicted (black traces) model outputs for simulated data using 3 of 20 available input signals. (A) Results when the optimal inputs are chosen and the robust identification algorithm is used. (B) Results when the inputs are randomly chosen and a pseudo-inverse is not used in the identification process. Simulation parameters: $K = 10$, $N = 20$, 4000 data points for each of the simulated data vectors (2000 points were used for the estimation and 2000 for the cross-validation). The estimated linear filters had a length of $M = 32$ points.

identification, as compared to the result obtained using the original identification algorithm. This increase corresponds to a nearly twofold reduction in the mean squared error. In a series of 100 Monte Carlo simulations, r^2 increased by 0.17 ± 0.11 .

4.2 System Identification from Physiological Data. The algorithms presented in this article were designed to work with correlated multiple input data of restricted bandwidth. The experimental data collected from the microelectrode arrays exhibited both of these features. Figure 6A shows

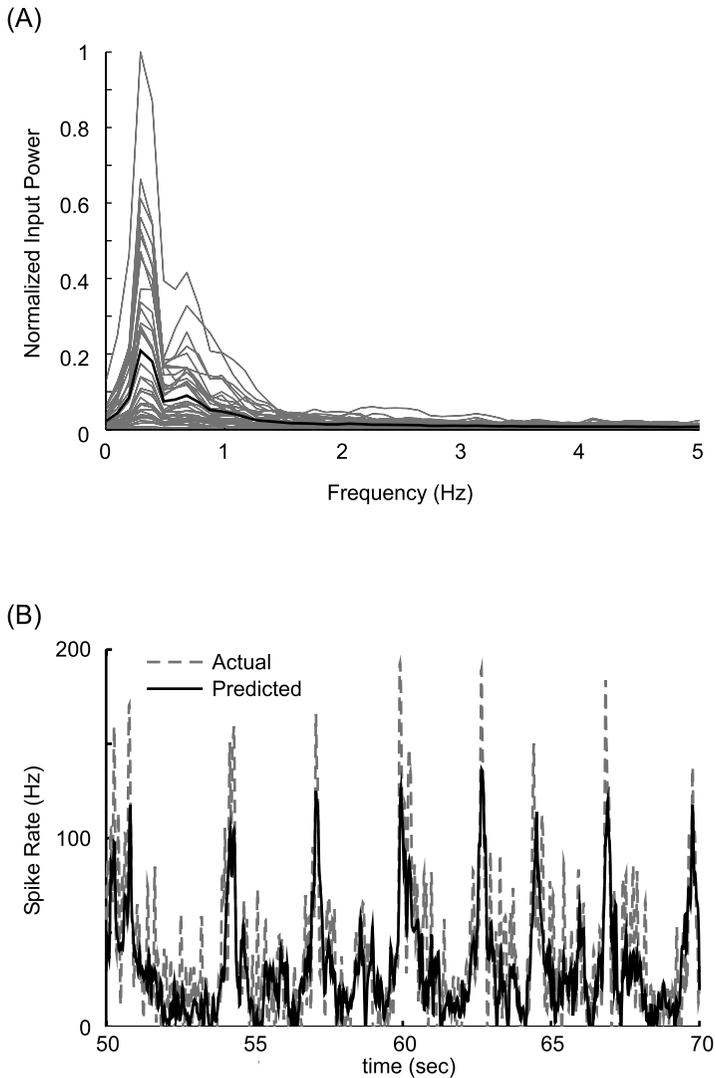


Figure 6: Input data characteristics. (A) Power spectra for the cortical recordings for a typical 180 second reaching trial. Individual signals are shown in gray and the mean of all signals in black. (B) Interdependence of the cortical recordings. The gray trace shows the instantaneous spike rate recorded from a single electrode; the black trace shows the estimate of that recording based on the recordings from all other electrodes (39 signals). Both signals have been low-pass-filtered at 10 Hz using a fourth-order Butterworth filter. The prediction is based on a linear MISO system estimated from 180 seconds of data. All estimated filters had a length of 510 ms.

the power spectra for all cortical signals recorded during a typical trial; the average spectrum is shown in black. The spectrum for each signal was broad but had a dominant peak at approximately 0.2 Hz to 0.5 Hz, due to the time between subsequent reaching movements. The power outside this range was lower but remained significant because the signals contained relatively narrow bursts of activity. Correlation between channels was assessed by examining the accuracy with which a given neural signal could be predicted using all other available neural signals. A causal linear MISO system was used for this prediction. Across all 29 data sets, the squared correlation coefficient between any given input and its prediction from all of the other inputs was 0.33 ± 0.10 (mean \pm SD). Furthermore, the best input prediction in each trial resulted in r^2 of 0.55 ± 0.03 . Figure 6B shows a typical example of the dependence between recorded neural signals. The gray trace shows the spike rate of the neural signal recorded from a single electrode, and the black trace shows the prediction of that signal based on the measurements from all other electrodes. Both signals were low-pass-filtered at 10 Hz to accentuate the average spike rates. The r^2 value for these two signals prior to filtering was 0.32. These results demonstrate that the neural recordings obtained from the microelectrode array were not statistically independent and that there is significant overlap in the information contained in these recordings, especially at the lower frequencies more relevant to movement control.

The optimal selection algorithm was effective at reducing the number of cortical signals necessary to predict the EMG activity for each of the four recorded muscles. Figure 7 shows the average r^2 value across all 29 data sets as a function of the number of inputs used to predict each of the recorded EMGs. Filter lengths of 510 ms were used for prediction; lengths beyond this size did not improve prediction accuracy significantly. The average r^2 obtained when using the optimally selected inputs is compared to that obtained when using the three alternative strategies: PCA, fSISO, and fMISO. The thick straight lines above each set of curves indicate the regions where the selection algorithm proposed in this letter performed significantly better than the alternatives ($p < 0.05$). The optimal algorithm produced the maximal cross-validation prediction accuracy for each of the four muscles tested and performed significantly better than the alternative methodologies. These results were most dramatic in comparison to the fSISO and PCA algorithms. There also was a statistically significant improvement in comparison to the fMISO algorithm, but the magnitude of the difference between the performance of that approach and that of the optimal algorithm was small.

Reducing the number of inputs also increased the accuracy of the EMG predictions for cross-validation data for 113 of the 116 available data sets (4 muscles \times 29 trials). The improvement in r^2 across all trials relative to when all inputs were used ranged from 0 to 0.28, with an average of 0.05 ± 0.07 . Because the number of optimal inputs varied between

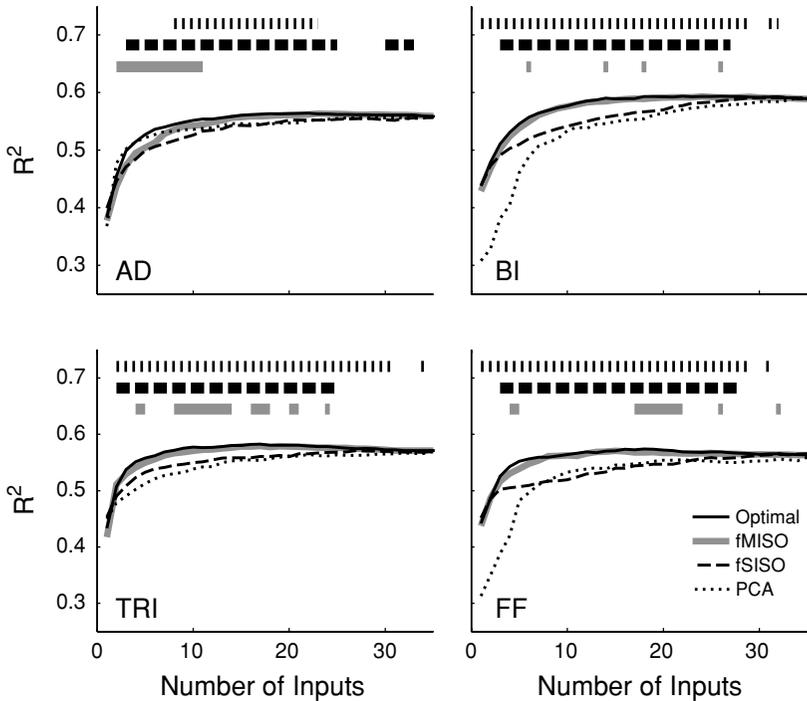


Figure 7: Performance of the optimal selection algorithm on electrophysiological data. Each panel corresponds to a different muscle and shows the average (29 trials) accuracy of EMG predictions as a function of the number of input signals used in the identification process. All reported r^2 values are for cross-validation data not used in the selection or identification processes. Solid traces correspond to r^2 obtained using the optimal selection algorithm. The gray lines show the results of using the fMISO selection algorithm, the coarse dashed lines show the results of using the fSISO selection algorithm, and the fine dashed lines show the results of using a PCA. In the estimation process, 120 seconds of data were used, and 60 seconds were used for cross-validation. All estimated filters had a length of 510 ms. The thick lines above the traces indicate regions where the optimal selection algorithm was significantly better ($p < 0.05$) than the alternative approaches.

trials, this improvement is not evident in the average responses shown in Figure 7.

Once the optimal set of inputs has been selected, the robust identification algorithm provided little performance enhancement with respect to techniques not based on a pseudo-inverse, as long as sufficient data were used in the identification process. Figure 8 provides results from a typical trial and shows r^2 for the fitted and cross-validation data, as a function of

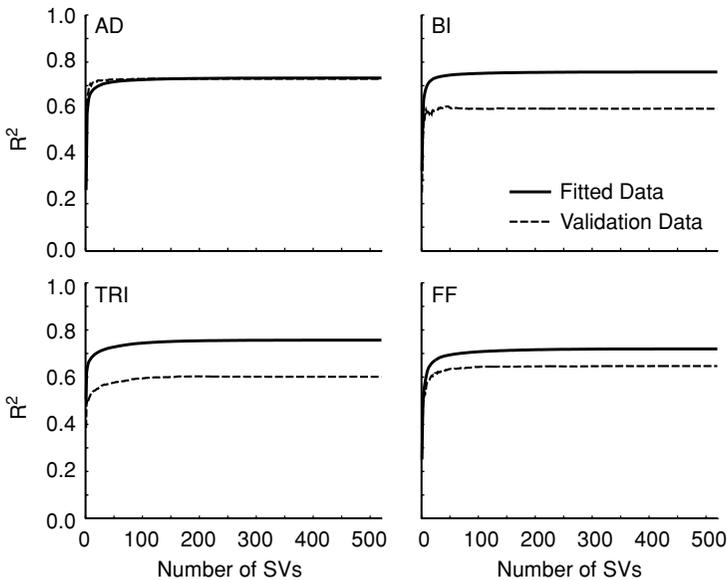


Figure 8: EMG prediction accuracy for the fitted (solid traces) and cross-validation (dashed traces) data as a function of the number of singular values used in the pseudo-inverse. For estimation, 120 seconds of data and 60 seconds for cross-validation were used. The estimated filters had a length of 510 ms.

the number of singular values used in the pseudo-inverse. This identification was performed on the 10 optimal inputs. Even when using these, less than 10% of the available singular values was needed to obtain 90% of the maximum achievable prediction accuracy. In contrast to the simulations presented in Figure 4, there is no clear peak in the cross-validation curve. This is due to the use of sufficiently long data records (2 minutes). The SVD algorithm had a more dramatic effect when fewer data were used in the identification process, although the peak value of r^2 was maximized by using at least 2 minutes of data for system identification.

The optimal algorithm for the selection of inputs has made it possible to predict upper limb EMGs reliably. Figure 9 compares the recorded and predicted EMGs from a set of cross-validation data for each muscle. The 10 best neurons were used for prediction; a different set of neurons was used for each muscle. There is a close correspondence between the actual and predicted EMGs. For the cross-validation data in these examples, r^2 was between 0.60 and 0.73, typical values for trials without long rest periods between subsequent movements.

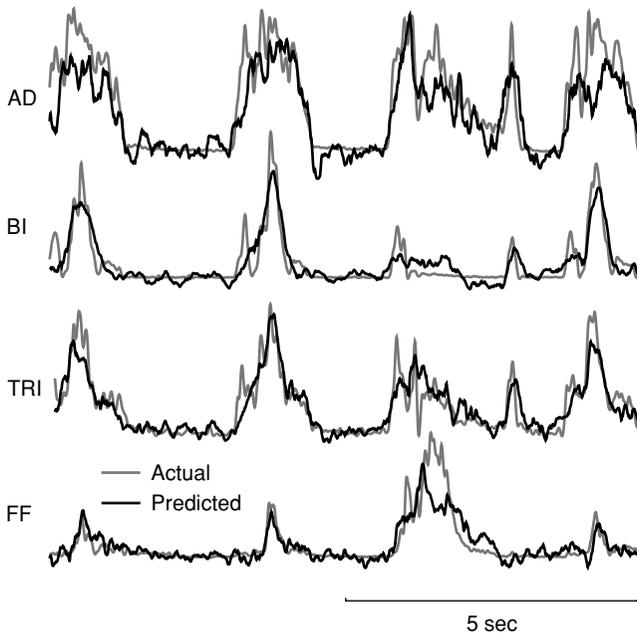


Figure 9: Actual (gray traces) and predicted (black traces) EMGs using 10 of 41 available cortical recordings. All plots are for cross-validation data not used in the estimation process. Models were estimated from 120 seconds of data, leaving 60 seconds available for cross-validation. The estimated filters had a length of 510 ms.

5 Discussion

In this letter, we have presented two novel algorithms to address numerical problems associated with the identification of MISO systems and have demonstrated the applicability of these algorithms to the processing of neural recordings from intracortical microelectrode arrays. These algorithms address the problems associated with identifying MISO systems with highly correlated inputs and inputs with restricted information content. The algorithms provide tools for selecting the optimal inputs to be used in the identification process and for generating robust estimates of the MISO system that relate these inputs to the observed output. Both algorithms were found to perform well for simulated and experimental data.

5.1 Selection of Inputs. With rapid advances in microelectrode technology, it becomes increasingly possible to obtain large numbers of simultaneous neural recordings. However, the computational burden associated with processing such large numbers of available inputs can be a challenge

in applications where efficient processing is essential. Furthermore, the likelihood for correlations among inputs increases as the number of recorded signals increases, as demonstrated in section 4. Such correlations can lead to numerical instabilities during the identification process (Perreault et al., 1999). Both issues can be addressed by eliminating inputs that do not provide unique information about the system output. Nonessential inputs may be uncorrelated with the output or may be highly correlated with other inputs. We have developed an algorithm for detecting such inputs and have demonstrated that it is effective. In addition to decreasing the computational time required for the remaining steps in the system identification process, such a pruning of neural inputs also has the potential to reduce the computational costs associated with preprocessing algorithms such as spike sorting and artifact removal.

The reported advantages associated with reducing the number of inputs used in the identification process are not in contradiction to findings that prediction accuracy increases with increasing numbers of recorded signals (Carmena et al., 2003; Wessberg et al., 2000). Recently, (Paninski et al., 2004) and (Sanchez et al., 2004) demonstrated that the accuracy with which movement variables can be predicted from neural recordings depends strongly on which neurons are selected as model inputs. Our results demonstrate that this selection can be optimized by choosing neural signals based on the uniqueness of their contribution to the system output. Increasing the number of neural recordings increases the sample of neurons from which to draw the optimal set. Therefore, the potential of experimental techniques that allow such large-scale recordings is likely to be enhanced by optimally selecting a subset of the available recorded signals for use in the identification and prediction process.

Similar selection algorithms recently were explored by Sanchez et al. (2004), who also demonstrated that a subset of the available neural recordings could be used to predict kinematic variables associated with reaching. We were able to compare two of their algorithms with the one proposed in this article. Although our selection algorithm produced the best results, the fMISO algorithm proposed by Sanchez et al. (2004) performed nearly as well. The results of both studies emphasize the need for considering all neural inputs and their contribution to the system output during the selection process.

5.2 Robust SVD Estimation. Most system identification algorithms rely on the use of white or at least broadband stationary inputs to produce reliable, robust estimates of the system dynamics. However, it can be difficult to obtain broadband inputs during functional behaviors. Under realistic conditions, the input bandwidth may be limited, and the assumption of stationarity may be violated. Hence, it is necessary to develop and use system identification algorithms that produce robust estimates of the system dynamics under such conditions. Westwick and Kearney

(1997b) have developed a robust algorithm for identifying SISO systems using a pseudo-inversion of the input autocorrelation matrix (see equation 2.8). Here we have extended this algorithm to the identification of MISO systems. Our results (see Figure 4) demonstrate that the use of this algorithm can improve prediction accuracy for cross-validation data not used in the identification process. These improvements are greatest when the number of data used in the identification process is relatively small, indicating that the robust algorithm helps reduce the problem of overfitting. Smaller improvements are observed when the number of data used to identify the MISO system is increased (see Figure 8). For neural data similar to those used in this study, this algorithm is likely to be most beneficial in situations where only small data records can be collected or when it is necessary to characterize system behavior over short periods of time.

5.3 Linear Systems Identification. This study has been restricted to the use of linear system identification techniques. Although the transformation from neural activity to motor output presumably contains many significant nonlinearities, linear models of the net transformation from neural activity to EMGs work surprisingly well when populations of neurons are considered. Similar phenomena have been demonstrated by a number of groups that have compared prediction accuracies of both linear filters and nonlinear networks for decoding neural information from motor and visual systems (Warland, Reinagel, & Meister, 1997; Wessberg et al., 2000; Gao et al., 2003). Given the similarities in prediction accuracy, there are a number of advantages to using linear identification techniques, including the computational and conceptual simplicity of these approaches as well as the potential for meaningful interpretation of the estimated linear filters. Linear IRFs can provide useful characterizations of the transfer of information from the cortex to the motor system (e.g., bandwidth, delays). In contrast, it can be more difficult to obtain similar insights if the system is modeled as a nonlinear network. However, the potential advantages of nonlinear models may become more apparent when their ability to generalize is tested under a wider range of conditions. For example, in situations where there is a significant pause between movements, it may be advantageous to incorporate a static output nonlinearity to approximate the threshold response of the motoneuron pool. Additionally, the techniques presented here could be applied to nonlinear systems consisting of static nonlinearities connected in series with finite memory linear systems (Bussgang, 1952). However, the applicability of these techniques to more general nonlinear systems remains to be demonstrated.

5.4 Potential Applications. The algorithms presented in this letter could be useful in a range of applications where it is necessary to predict the output of a MISO system and the inputs to that system are highly

correlated, have limited information content, or can be recorded for only short periods of time. With respect to the processing of neural information, the algorithms could be used in the analysis of multidimensional signals from a variety of sources, including electromyograms, electroencephalograms, and intracortical recordings.

One application for the input selection algorithm is as a mapping tool for determining which neural recordings are most relevant to any time-dependent process of interest. Examples include assessing the neural signals or anatomical substrates contributing to movement control, cognitive tasks, or visual processing. The selection algorithm also could be used in conjunction with the robust identification algorithm when it is necessary to predict the system output or generate a control signal from a set of recorded inputs. One such application that has received much recent attention is the development of BMIs, including those for the restoration of motor function via neuromuscular stimulation of paralyzed muscles (Lauer, Peckham, Kilgore, & Heetderks, 2000), the control of augmented communication aids for individuals with severe communication disorders (Kennedy, Bakay, Moore, Adams, & Goldwaithe, 2000), and the control of assistive devices for improved mobility and function (Carmena et al., 2003; Taylor, Tillery, & Schwartz, 2003). Success in each of these applications hinges on the availability of a multidimensional natural control signal, such as would result from intracortical microelectrode array recordings. The input selection algorithm could be used to identify the neural signals most relevant to each degree of control, and the robust identification algorithm could be used to estimate the system describing the dynamic relationship between those neural signals and the desired control signal. By using only the optimal inputs, it would be possible to decrease the computational time needed for identification and prediction. This could offer significant improvements in real-time applications, especially those using adaptive algorithms. It should be noted, though, that changes in the neural signals available over time will require a reevaluation of which of the available signals are optimal for a given task. This evaluation could operate as a background process, updating the set of optimal signals as necessary.

Acknowledgments

This research was supported by NSERC grant RGP-238939 (D.T.W.), NSF grant IBN-0432171 (E.J.P.) and NIH grants NS36976 (L.E.M.) and 1 K25 HD044720-01 (E.J.P.). E.A.P. was supported by NSF through an IGERT fellowship in Dynamics of Complex Systems in Science and Engineering, grant DGE-9987577. S.A.S. acknowledges the hospitality of the Kavli Institute for Theoretical Physics at the University of California, Santa Barbara, and partial NSF support under grant PHY99-07949.

References

- Bussgang, J. J. (1952). Crosscorrelation functions of amplitude distorted gaussian signals. *MIT Res. Lab. Elec. Tech. Rep.*, 216, 1–14.
- Carmena, J. M., Lebedev, M. A., Crist, R. E., O'Doherty, J. E., Santucci, D. M., Dimitrov, D., Patil, P. G., Henriquez, C. S., & Nicolelis, M. A. (2003). Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol.*, 1(2), E42.
- Chapin, J. K., Moxon, K. A., Markowitz, R. S., & Nicolelis, M. A. (1999). Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nat. Neurosci.*, 2(7), 664–670.
- Chen, S., Cowan, C., & Grant, P. (1991). Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Trans. Neural Netw.*, 2, 302–309.
- Donoghue, J. P. (2002). Connecting cortex to machines: Recent advances in brain interfaces. *Nat. Neurosci.*, 5 (Suppl.), 1085–1088.
- Gao, Y., Black, M. J., Bienenstock, E., Wu, W., & Donoghue, J. P. (2003). A quantitative comparison of linear and non-linear models of motor cortical activity for the encoding and decoding of arm motions. In *1st International IEEE/EMBS Conference on Neural Engineering* (pp. 189–192). Los Alamitos, CA: IEEE.
- Golub, G., & Van Loan, C. (1989). *Matrix computations* (2nd ed.). Baltimore, MD: Johns Hopkins University Press.
- Isaacs, R. E., Weber, D. J., & Schwartz, A. B. (2000). Work toward real-time control of a cortical neural prosthesis. *IEEE Trans. Rehabil. Eng.*, 8(2), 196–198.
- Kennedy, P. R., Bakay, R. A., Moore, M. M., Adams, K., & Goldwithe, J. (2000). Direct control of a computer from the human central nervous system. *IEEE Trans. Rehabil. Eng.*, 8(2), 198–202.
- Kim, S. P., Sanchez, J. C., Erdogmus, D., Rao, Y. N., Wessberg, J., Principe, J. C., & Nicolelis, M. (2003). Divide-and-conquer approach for brain machine interfaces: Nonlinear mixture of competitive linear models. *Neural Netw.*, 16(5–6), 865–871.
- Korenberg, M. (1988). Identifying nonlinear difference equation and functional expansion representations: The fast orthogonal algorithm. *Ann. Biomed. Eng.*, 16, 123–142.
- Lauer, R. T., Peckham, P. H., Kilgore, K. L., & Heetderks, W. J. (2000). Applications of cortical signals to neuroprosthetic control: A critical review. *IEEE Trans. Rehabil. Eng.*, 8(2), 205–208.
- Maynard, E. M., Nordhausen, C. T., & Normann, R. A. (1997). The Utah Intracortical Electrode Array: A recording structure for potential brain-computer interfaces. *Electroencephalogr. Clin. Neurophysiol.*, 102(3), 228–239.
- Miller, A. (1990). *Subset selection in regression*. London: Chapman and Hall.
- Mussa-Ivaldi, F. A., & Miller, L. E. (2003). Brain-machine interfaces: Computational demands and clinical needs meet basic neuroscience. *Trends Neurosci.*, 26(6), 329–334.
- Nicolelis, M. A., Dimitrov, D., Carmena, J. M., Crist, R., Lehw, G., Kralik, J. D., & Wise, S. P. (2003). Chronic, multisite, multielectrode recordings in macaque monkeys. *Proc. Natl. Acad. Sci. USA*, 100(19), 11041–11046.

- Paninski, L., Fellows, M. R., Hatsopoulos, N. G., & Donoghue, J. P. (2004). Spatiotemporal tuning of motor cortical neurons for hand position and velocity. *J. Neurophysiol.*, *91*(1), 515–532.
- Perreault, E., Kirsch, R., & Acosta, A. (1999). Multiple-input, multiple-output system identification for characterization of limb stiffness dynamics. *Biol. Cybern.*, *80*, 327–337.
- Sanchez, J., Carmena, J., Lebedev, M., Nicolelis, M., Harris, J., & Principe, J. (2004). Ascertaining the importance of neurons to develop better brain-machine interfaces. *IEEE Trans. Biomed. Eng.*, *51*(6), 943–953.
- Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R., & Donoghue, J. P. (2002). Instant neural control of a movement signal. *Nature*, *416*(6877), 141–142.
- Taylor, D. M., Tillery, S. I., & Schwartz, A. B. (2002). Direct cortical control of 3D neuroprosthetic devices. *Science*, *296*(5574), 1829–1832.
- Taylor, D. M., Tillery, S. I., & Schwartz, A. B. (2003). Information conveyed through brain-control: Cursor versus robot. *IEEE Trans. Neural. Syst. Rehabil. Eng.*, *11*(2), 195–199.
- Warland, D. K., Reinagel, P., & Meister, M. (1997). Decoding visual information from a population of retinal ganglion cells. *J. Neurophysiol.*, *78*(5), 2336–2350.
- Wessberg, J., Stambaugh, C. R., Kralik, J. D., Beck, P. D., Laubach, M., Chapin, J. K., Kim, J., Biggs, S. J., Srinivasan, M. A., & Nicolelis, M. A. (2000). Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, *408*(6810), 361–365.
- Westwick, D. T., & Kearney, R. E. (1997a). Generalized eigenvector algorithm for nonlinear system identification with non-white inputs. *Ann. Biomed. Eng.*, *25*(5), 802–814.
- Westwick, D. T., & Kearney, R. (1997b). Identification of physiological systems: A robust method for non-parametric impulse response estimation. *Med. Biol. Eng. Comput.*, *35*(2), 83–90.
- Williams, J. C., Rennaker, R. L., & Kipke, D. R. (1999). Long-term neural recording characteristics of wire microelectrode arrays implanted in cerebral cortex. *Brain Res. Protoc.*, *4*(3), 303–313.
- Wu, W., Black, M. J., Mumford, D., Gao, Y., Bienenstock, E., & Donoghue, J. (2003). A switching Kalman filter model for the motor cortical coding of hand motion. In *25th International IEEE/EMBS Conference*. Los Alamitos, CA: IEEE.