Cosine Tuning Minimizes Motor Errors

Emanuel Todorov
emo@gatsby.ucl.ac.uk
Gatsby Computational Neuroscience Unit, University College London, London, U.K.

Cosine tuning is ubiquitous in the motor system, yet a satisfying explanation of its origin is lacking. Here we argue that cosine tuning minimizes expected errors in force production, which makes it a natural choice for activating muscles and neurons in the final stages of motor processing. Our results are based on the empirically observed scaling of neuromotor noise, whose standard deviation is a linear function of the mean. Such scaling predicts a reduction of net force errors when redundant actuators pull in the same direction. We confirm this prediction by comparing forces produced with one versus two hands and generalize it across directions. Under the resulting neuromotor noise model, we prove that the optimal activation profile is a (possibly truncated) cosine—for arbitrary dimensionality of the workspace, distribution of force directions, correlated or uncorrelated noise, with or without a separate cocontraction command. The model predicts a negative force bias, truncated cosine tuning at low muscle cocontraction levels, and misalignment of preferred directions and lines of action for nonuniform muscle distributions. All predictions are supported by experimental data.

1 Introduction

Neurons are commonly characterized by their tuning curves, which describe the average firing rate $f(x)$ as a function of some externally defined variable x. The question of what constitutes an optimal tuning curve for a population code (Hinton, McClelland, & Rumelhart, 1986) has attracted considerable attention. In the motor system, cosine tuning has been well established for motor cortical cells (Georgopoulos, Kalaska, Caminiti, & Massey, 1982; Kettner, Schwartz, & Georgopoulos, 1988; Kalaska, Cohen, Hyde, & Prud'homme, 1989; Caminiti, Johnson, Galli, Ferraina, & Burnod, 1991) as well as individual muscles (Turner, Owens, & Anderson, 1995; Herrmann & Flanders, 1998; Hoffman & Strick, 1999).[1] The robustness of cosine

---

[1] When a subject exerts isometric force or produces movements, each cell and muscle is maximally active for a particular direction of force or movement (called *preferred direction*), and its activity falls off with the cosine of the angle between the preferred and actual direction.

tuning suggests that it must be optimal in some meaningful sense, yet a satisfactory explanation is lacking.

In this article, we argue that cosine tuning in the motor system is indeed optimal in the most meaningful sense one can imagine: it minimizes the net effect of neuromotor noise, resulting in minimal motor errors. The argument developed here is specific to the motor system. Since it deviates from previous analyses of optimal tuning, we begin by clarifying the main differences.

1.1 Alternative Approaches to Optimal Tuning. The usual approach (Hinton et al., 1986; Snippe, 1996; Zhang & Sejnowski, 1999b; Pouget, Deneve, Ducom, & Latham, 1999) is to equate the goodness of a tuning function f with how accurately the variable x can be reconstructed from a population of noisy responses $\mu_1 + \varepsilon_1, \ldots, \mu_n + \varepsilon_n$, where $\mu_i = f(x - c_i)$ is the mean response of neuron i with receptive field center c_i. This approach to the analysis of empirically observed tuning is mathematically appealing but involves hard-to-justify assumptions:

- In the absence of data on higher-order correlations and in the interest of analytical tractability, oversimplified noise models have to be assumed.[2] In contrast, when the population responses are themselves the outputs of a recurrent network, the noise joint distribution is likely to be rather complex. Ignoring that complexity can lead to absurd conclusions, such as an apparent increase of information (Pouget et al., 1999).

- Since the actual reconstruction mechanisms used by the nervous system as well as their outputs are rarely observable, one has to rely on theoretical limits (i.e., the Cramér-Rao bound), ignoring possible biological constraints and noise originating at the reconstruction stage. Optimality criteria that may arise from the need to perform computation (and not just represent or transmit information) are also ignored.

Even if these assumptions are accepted, it was recently shown (Zhang & Sejnowski, 1999b) that the optimal tuning width is biologically implausible: as narrow as possible[3] when x is one-dimensional, irrelevant when x is two-dimensional, and as broad as possible when x is more than two-dimensional. Thus, empirical observations such as cosine tuning are difficult to interpret as being optimal in the usual sense.

---

[2] The noise terms $\varepsilon_1, \ldots, \varepsilon_n$ are usually modeled as independent or homogeneously correlated Poisson variables.

[3] The finite number of neurons in the population prevents infinitely sharp tuning (i.e., the entire range of x has to be covered), but that is a weak constraint since a given area typically contains large numbers of neurons.

In this article, we pursue an alternative approach. The optimal tuning function f^* is still defined as the one that maximizes the accuracy of the reconstruction $\mu_1 + \varepsilon_1, \ldots, \mu_n + \varepsilon_n \to \widehat{x}$. However, we do not speculate that the input noise distribution has any particular form or that the reconstruction is optimal. Instead, we use knowledge of the actual reconstruction mechanisms and measurements of the actual output $\widehat{x}$, which in the motor system is simply the net muscle force.[4] That allows us to infer a direct mapping $\mu_1, \ldots, \mu_n \to \text{Mean}(\widehat{x}), \text{Var}(\widehat{x})$ from the mean of the inputs to the mean and variance of the output. Once such a mapping is available, the form of the input noise and the amount of information about x that in principle could have been extracted become irrelevant to the investigation of optimal tuning.

1.2 Optimal Tuning in the Motor System. We construct the mapping $\mu_1, \ldots, \mu_n \to \text{Mean}(\widehat{x}), \text{Var}(\widehat{x})$ based on two sets of observations, relating (1) the mean activations to the mean of the net force and (2) the mean to the variance of the net force.

Under isometric conditions, individual muscles produce forces in proportion to the rectified and filtered electromyogram (EMG) signals (Zajac, 1989; Winter, 1990), and these forces add mechanically to the net force.[5] Thus, the mean of the net force is simply the vector sum of the mean muscle activations $\mu_1, \ldots, \mu_n$ multiplied by the corresponding force vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ (defining the constant lines of action). If the output cells of primary motor cortex (M1) contribute additively to the activation of muscle groups (Todorov, 2000), a similar additive model may apply for $\mu_1, \ldots, \mu_n$ corresponding to mean firing rates in M1. In the rest of the article, $\mu_1, \ldots, \mu_n$ will denote the mean activation levels of abstract force generators, which correspond to individual muscles or muscle groups. The relevance to M1 cell tuning is addressed in Section 6.

Numerous studies of motor tremor have established that the standard deviation of the net force increases linearly with its mean. This has been demonstrated when tonic isometric force is generated by muscle groups (Sutton & Sykes, 1967) or individual muscles (McAuley, Rothwell, & Marsden, 1997). The same scaling holds for the magnitude of brief force pulses (Schmidt, Zelaznik, Hawkins, Frank, & Quinn, 1979). This general finding

---

[4] We focus predominantly on isometric force tasks and extend our results to movement velocity and displacement tuning in the last section. Thus, the output (reconstruction) is defined as net force (i.e., vector sum of all individual muscle forces) in the relevant work space.

[5] The contributions of different muscles to end-point force are determined by the Jacobian transformation and the tendon insertion points. Each muscle has a line of action (force vector) in end-point space, as well as in joint space. Varying the activation level under isometric conditions affects the force magnitude, but the force direction remains fixed.

is also confirmed indirectly by the EMG histograms of various muscles, which lie between a gaussian and a Laplace distribution, both centered at 0 (Clancy & Hogan, 1999). Under either distribution, the rectified signal $|x|$ has standard deviation proportional to the mean.[6]

The above scaling law leads to a neuromotor noise model where each generator contributes force with standard deviation σ linear in the mean μ: $\sigma = a\mu$. This has an interesting consequence. Suppose we had two redundant generators pulling in the same direction and wanted them to produce net force μ. If we activated only one of them at level μ, the net variance would be $\sigma^2 = a^2\mu^2$. If we activated both generators at level $\mu/2$, the net variance (assuming uncorrelated noise) would be $\sigma^2 = a^2\mu^2/2$, which is two times smaller. Thus, it is advantageous to activate all generators pulling in the direction of desired net force. What about generators pulling in slightly different directions? If all of them are recruited simultaneously, the noise in the net force direction will still decrease, but at the same time, extra noise will be generated in orthogonal directions. So the advantage of activating redundant actuators decreases with the angle away from the net force direction. The main technical contribution of this article is to show that it decreases as a cosine, that is, cosine tuning minimizes expected motor errors.

Note that the above setting of the optimal tuning problem is in open loop; the effects of activation level on feedback gains are not explicitly considered. Such effects should be taken into account because coactivation of opposing muscles may involve interesting trade-offs: it increases both neuromotor noise and system impedance and possibly modifies sensory inputs (due to $\alpha - \gamma$ coactivation). We incorporate these possibilities by assuming that an independent cocontraction command C may be specified, in which case the net activity of all generators is constrained to be equal to C. As shown below, the optimal tuning curve is a cosine regardless of whether C is specified. The optimal setting of C itself will be addressed elsewhere.

In the next section, we present new experimental results, confirming the reduction of noise due to redundancy. The rest of the article develops the mathematical argument for cosine tuning rigorously, under quite general assumptions.

2 Actuator Redundancy Decreases Neuromotor Noise

The empirically observed scaling law $\sigma = a\mu$ implies that activating redundant actuators should reduce the overall noise level. This effect forms the basis of the entire model, so we decided to test it experimentally. Ideally, we would ask subjects to produce specified forces by activating one versus two

---

[6] For the Laplace distribution $p_\sigma(x) = \frac{1}{\sigma}\exp(\frac{-|x|}{\sigma})$, the mean of $|x|$ is σ and the variance is σ^2. For the 0-mean gaussian with standard deviation σ, the mean of $|x|$ is $\sigma\sqrt{2/\pi}$, and the variance is $\sigma^2(1 - 2/\pi)$.

synergistic muscles and compare the corresponding noise levels. But human subjects have little voluntary control over which muscles they activate, so instead we used the two hands as redundant force generators: we compared the force errors for the same level of net instructed force produced with one hand versus both hands. The two comparisons are not identical, since the neural mechanisms coordinating the two hands may be different from those coordinating synergistic muscles of one limb. In particular, one might expect coordinating the musculature of both hands to be more difficult, which would increase the errors in the two-hands condition (opposite to our prediction). Thus, we view the results presented here as strong supporting evidence for the predicted effect of redundancy on neuromotor noise.

2.1 Methods. Eight subjects produced isometric forces of specified magnitude (3–33 N) by grasping a force transducer disk (Assurance Technologies F/T Gamma 65/5, 500 Hz sampling rate, 0.05 N resolution) between the thumb and the other four fingers. The instantaneous force magnitude produced by the subject was displayed with minimum delay as a vertical bar on a linear 0–40N scale. Each of 11 target magnitudes was presented in a block of three trials (5 sec per trial, 2 sec between trials), and the subjects were asked to maintain the specified force as accurately as possible. The experiment was repeated twice: with the dominant hand and with both hands grasping the force transducer. Since forces were measured along the forward axis, the two hands can be considered as mechanically identical (i.e., redundant) actuators. To balance possible learning and fatigue effects, the order of the 11 force magnitudes was randomized separately for each subject (subsequent analysis revealed no learning effects). Half of the subjects started with both hands, the other half with the dominant hand. The first 2 seconds of each trial were discarded; visual inspection confirmed that the 2 second initial period contained the force transient associated with reaching the desired force level. The remaining 3 seconds (1500 sample points) of each trial were used in the data analysis.

2.2 Results. The average standard deviations are shown in Figure 1B for each force level and hand condition. In agreement with previous results, the standard deviation in both conditions was a convincingly linear function of the instructed force level. As predicted, the force errors in the two-hands condition were smaller, and the ratio of the two slopes was 1.42 ± 0.25 (95% confidence interval), which is indistinguishable from the predicted value of $\sqrt{2} \approx 1.41$. Two-way (2 conditions $\times$ 11 force levels) ANOVA with replications (eight subjects) indicated that both effects were highly significant ($p < 0.0001$), and there was no interaction effect ($p = 0.57$). Plotting standard deviation versus mean (rather than instructed) force produced very similar results.
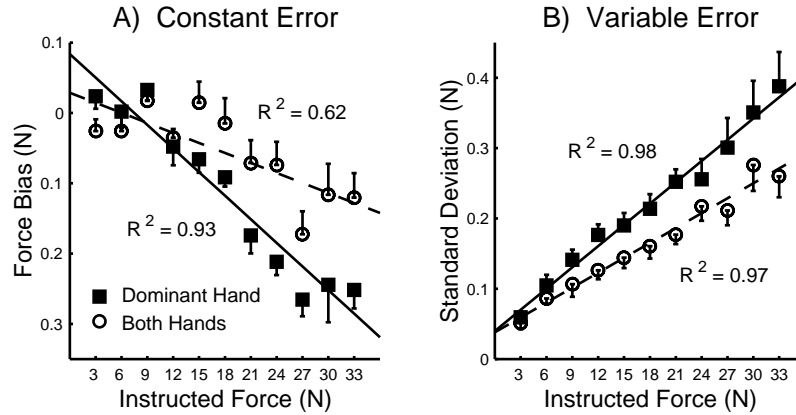
Figure 1: The last 3 seconds of each trial were used to estimate the bias (A) and standard deviation (B) for each instructed force level and hand condition. Averages over subjects and trials, with standard error bars, are shown in the figure. The standard deviation estimates were corrected for sensor noise, measured by placing a 2.5 kg object on the sensor and recording for 10 seconds.

The nonzero intercept in our data was smaller than previous observations but still significant. It is not due to sensor noise (as previously suggested), because we measured that noise and subtracted its variance. One possible explanation is that because of cocontraction, some force fluctuations are present even when the mean force is 0.

Figure 2A shows the power spectral density of the fluctuations in the two conditions, separated into low (3–15N) and high (21–33N) force levels. The scaling is present at all frequencies, as observed previously (Sutton & Sykes, 1967). Both increasing actuator redundancy and decreasing the force level have similar effects on the spectral density. To identify possible differences between frequency bands, we low-pass-filtered the data at 5 Hz, and bandpass filtered at 5–25 Hz. As shown in Figure 2B, the noise in both frequency bands obeys the same scaling law: standard deviation linear in the mean, with higher slope in the one-hand condition. The slopes in the two frequency bands are different, and interestingly the intercept we saw before is restricted to low frequencies. If the nonzero intercept is indeed due to cocontraction, Figure 2B implies that the cocontraction signal (i.e., common central input to opposing muscles) fluctuates at low frequencies.

We also found small but highly significant negative biases (defined as the difference between measured and instructed force) that increased with the instructed force level and were higher in the one-hand condition (see Figure 1A). This effect cannot be explained with perceptual or memory
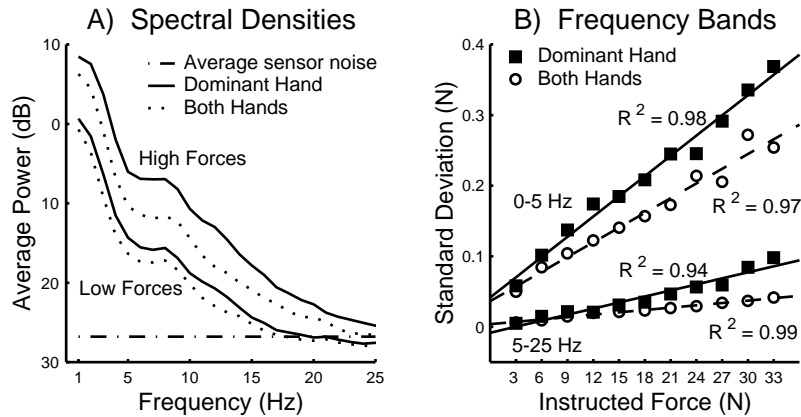
Figure 2: (A) The power spectral density of the force fluctuations was estimated using blocks of 500 sample points with 250 point overlap. Blocks were mean-detrended, windowed using a Hanning window, and the squared magnitude of the Fourier coefficients averaged separately for each frequency. This was done separately for each instructed force level and then the low (3–15 N) and high (21–33 N) force levels were averaged. The data from some subjects showed a sharper peak around 8–10 Hz, but that is smoothed out in the average plot. There appears to be a qualitative change in the way average power decreases at about 5 Hz. (B) The data were low-pass-filtered at 5 Hz (fifth-order Butterworh filter) and also bandpass filtered at 5–25 Hz. The standard deviation for each force level and hand condition was estimated separately in each frequency band.

limitations, since subjects received real-time visual feedback on a linear force scale. A similar effect is predicted by optimal force production: if the desired force level is μ^* and we specify mean activation μ for a single generator, the expected square error is $(\mu - \mu^*)^2 + a^2\mu^2$, which is minimal for $\mu = \mu^*/(1 + a^2) < \mu^*$. Thus, the optimal bias is negative, larger in the one-hand condition, and its magnitude increases with μ^*. The slopes in Figure 1A are substantially larger than predicted, which is most likely due to a trade-off between error and effort (see section 5.2). Other possible explanations include an inaccurate internal estimate of the noise magnitude and a cost function that penalizes large fluctuations more than the square error cost does (see section 5.4).

Summarizing the results of the experiment, the neuromotor noise scaling law observed previously (Sutton & Sykes, 1967; Schmidt et al., 1979) was replicated. Our prediction that redundant generators reduce noise was confirmed. Thus, we feel justified in assuming that each generator contributes force whose standard deviation is a linear function of the mean.

3 Force Production Model

Consider a set Ω of force generators producing force (torque) in D-dimensional Euclidean space $\mathbb{R}^D$. Generator $\alpha \in \Omega$ produces force proportional to its instantaneous activation, always in the direction of the unit vector $\mathbf{u}(\alpha) \in \mathbb{R}^D$.[7] The dimensionality D can be only 2 or 3 for end-point force but much higher for joint torque—for example, 7 for the human arm.

The central nervous system (CNS) specifies the mean activations $\mu(\alpha)$, which are always nonnegative. The actual force contributed by each generator is $(\mu(\alpha) + z(\alpha))\mathbf{u}(\alpha)$. The neuromotor noise z is a set of zero-mean random variables whose probability distribution $p(z|\mu)$ depends on μ (see below), and $p(z(\alpha) < -\mu(\alpha)) = 0$ since muscles cannot push. Thus, the net force $\mathbf{r}(\mu) \in \mathbb{R}^D$ is the random variable:[8]

$$\mathbf{r}(\mu) = |\Omega|^{-1} \sum_{\alpha \in \Omega} (\mu(\alpha) + z(\alpha))\mathbf{u}(\alpha).$$

Given a desired net force vector $\mathbf{f} \in \mathbb{R}^D$ and optionally a cocontraction command-net activation $C = |\Omega|^{-1} \sum_\alpha \mu(\alpha)$, the task of the CNS is to find the activation profile $\mu(\alpha) \geq 0$ that minimizes the expected force error under $p(z|\mu)$.[9] We will define error as the squared Euclidean distance between the desired force $\mathbf{f}$ and actual force $\mathbf{r}$ (alternative cost functions are considered in section 5.4). Note that both direction and magnitude errors are penalized, since both are important for achieving the desired motor objectives. The expected error is the sum of variance V and squared bias B,

$$\mathrm{E}_{z|\mu}\left[(\mathbf{r} - \mathbf{f})^T(\mathbf{r} - \mathbf{f})\right] = \underbrace{\mathrm{trace}(\mathrm{Cov}_{z|\mu}[\mathbf{r}, \mathbf{r}])}_{V} + \underbrace{(\bar{\mathbf{r}} - \mathbf{f})^T(\bar{\mathbf{r}} - \mathbf{f})}_{B}, \qquad (3.1)$$

where the mean force is $\bar{\mathbf{r}} = |\Omega|^{-1} \sum_\alpha \mu(\alpha)\mathbf{u}(\alpha)$ since $\mathrm{E}[z(\alpha)] = 0$ for each α.

We first focus on minimizing the variance term V for specified mean force $\bar{\mathbf{r}}$ and then perform another minimization with respect to (w.r.t.) $\bar{\mathbf{r}}$. Exchang-

---

[7] α will be used interchangeably as an index over force generators in the discrete case and as a continuous index specifying direction in the continuous case.

[8] The scaling constant $|\Omega|^{-1}$ simplifies the transition to a continuous Ω later: $|\Omega|^{-1} \sum_\alpha \cdots$ will be replaced with $|\mathcal{S}_D|^{-1} \int \cdots d\alpha$ where $|\mathcal{S}_D|$ is the surface area of the unit sphere in $\mathbb{R}^D$. It does not affect the results.

[9] $\mu(\alpha)$ is the activation profile of all generators at one point in time, corresponding to a given net force vector. In contrast, a tuning curve is the activation of a single generator when the net force direction varies. When $\mu(\alpha)$ is symmetric around the net force direction, it is identical to the tuning curve of all generators. This symmetry holds in most of our results, except for nonuniform distributions of force directions. In that case, we will compute the tuning curve explicitly (see Figure 4).

ing the order of the trace, $\mathrm{E}, \sum$ operators, the variance term becomes:

$$V = |\Omega|^{-2} \operatorname{trace}\left(\mathrm{E}_{z|\mu}\left[\sum_{\alpha\in\Omega}\sum_{\beta\in\Omega} z(\alpha)z(\beta)\mathbf{u}(\alpha)\mathbf{u}(\beta)^T\right]\right)$$

$$= |\Omega|^{-2}\sum_{\alpha\in\Omega}\sum_{\beta\in\Omega}\operatorname{Cov}_{z|\mu}[z(\alpha), z(\beta)]\mathbf{u}(\beta)^T\mathbf{u}(\alpha).$$

To evaluate V, we need a definition of the noise covariance $\operatorname{Cov}_{z|\mu}[z(\alpha), z(\beta)]$ for any pair of generators α, β. Available experimental results only suggest the form of the expression for $\alpha = \beta$; since the standard deviation is a linear function of the mean force, $\operatorname{Cov}_{z|\mu}[z(\alpha), z(\alpha)]$ is a quadratic polynomial of $\mu(\alpha)$. This will be generalized to a quadratic polynomial across directions as

$$\operatorname{Cov}_{z|\mu}[z(\alpha), z(\beta)] = \left(\lambda_1\mu(\alpha)\mu(\beta) + \lambda_2\frac{\mu(\alpha) + \mu(\beta)}{2}\right)(\delta_\alpha^\beta + \lambda_3).$$

The δ_α^β term is a delta function corresponding to independent noise for each force generator. The correlation term λ_3 corresponds to fluctuations in some shared input to all force generators. A correlation term dependent on the angle between $\mathbf{u}(\alpha)$ and $\mathbf{u}(\beta)$ is considered in section 5.3.

Substituting in the above expression for V and defining $\mathbf{U} = |\Omega|^{-1}\sum_\alpha \mathbf{u}(\alpha)$, which is $\mathbf{0}$ when the force directions are uniformly distributed, we obtain

$$V = |\Omega|^{-2}\sum_{\alpha\in\Omega}(\lambda_1\mu(\alpha)^2 + \lambda_2\mu(\alpha)) + \lambda_1\lambda_3\bar{\mathbf{r}}^T\bar{\mathbf{r}} + \lambda_2\lambda_3\bar{\mathbf{r}}^T\mathbf{U}. \qquad (3.2)$$

The optimal activation profile can be computed in two steps: (1) for given $\bar{\mathbf{r}}$ in equation 3.2, find the constrained minimum $V^*(\bar{\mathbf{r}})$ w.r.t. μ; (2) substitute in equation 3.1 and find the minimum of $V^*(\bar{\mathbf{r}}) + B(\bar{\mathbf{r}})$ w.r.t. $\bar{\mathbf{r}}$. Thus, the shape of the optimal activation profile emerges in step 1 (see section 4), while the optimal bias is found in step 2 (see section 5.1).

4 Cosine Tuning Minimizes Expected Motor Error

The minimization problem given by equation 3.2 is an instance of a more general minimization problem described next. We first solve that general problem and then specialize the solution to equation 3.2.

The set Ω we consider can be continuous or discrete. The activation function (vector) $\mu \in \mathcal{R}(\Omega)$: $\Omega \to \mathbb{R}$ is nonnegative for all $\alpha \in \Omega$. Given arbitrary positive weighting function $w \in \mathcal{R}(\Omega)$, projection functions $g_{1,...,N} \in \mathcal{R}(\Omega)$, resultant lengths $r_{1,...,N} \in \mathbb{R}$, and offset $\lambda \in \mathbb{R}$, we will solve the following

minimization problem w.r.t. μ:

$$
\begin{array}{ll}
\text{Minimize} & \langle \mu + \lambda, \mu + \lambda \rangle \\
\text{Subject to} & \mu(\alpha) \geq 0 \quad \langle \mu, g_{1...N} \rangle = r_{1,...,N} \\
\text{Where} & \langle u, v \rangle \triangleq |\Omega|^{-1} \sum_{\alpha \in \Omega} u(\alpha)v(\alpha)w(\alpha).
\end{array}
\tag{4.1}
$$

The generalized dot product is symmetric, linear in both arguments, and positive definite (since $w(\alpha) > 0$ by definition). Note that the dot product is defined between activation profiles rather than force vectors.

The solution to this general problem is given by the following result (see the appendix):[10]

Theorem 1.   *If $\mu^*(\alpha) = \lfloor \sum_n a_n g_n(\alpha) - \lambda \rfloor$ satisfies $\langle \mu^*, g_{1,...,N} \rangle = r_{1,...,N}$ for some $a_{1,...,N} \in \mathbb{R}$, then μ^* is the unique constrained minimum of $\langle \mu + \lambda, \mu + \lambda \rangle$.*

Thus, the unique optimal solution to equation 4.1 is a truncated linear combination of the projection functions $g_{1,...,N}$, assuming that a set of N constants $a_{1,...,N} \in \mathbb{R}$ satisfying the N constraints $\langle \mu^*, g_{1,...,N} \rangle = r_{1,...,N}$ exists. Although we have not been able to prove their existence in general, for the concrete problems of interest, these constants can be found by construction (see below). Note that the analytic form of μ^* does not depend on the arbitrary weighting function w used to define the dot product (the numerical values of the constants $a_{1,...,N}$ can of course depend on w).

In the case $\sum_n a_n g_n(\alpha) \geq \lambda$ for all $\alpha \in \Omega$, the constants $a_{1,...,N}$ satisfy the system of linear equations $\sum_n a_n \langle g_n, g_k \rangle = r_k + \langle \lambda, g_k \rangle$ for $k = 1, \ldots, N$. It can be solved by inverting the matrix $G_{nk} \triangleq \langle g_n, g_k \rangle$; for the functions $g_{1,...,N}$ considered in section 4.3, the matrix G is always invertible.

4.1 Application to Force Generation. We now clarify what this general result has to do with our problem. Recall that the goal is to find the nonnegative activation profile $\mu(\alpha) \geq 0$ that minimizes equation 3.2 for given net force $\bar{\mathbf{r}} = |\Omega|^{-1} \sum_\alpha \mu(\alpha)\mathbf{u}(\alpha)$ and optionally cocontraction $C = |\Omega|^{-1} \sum_\alpha \mu(\alpha)$. Omitting the last two terms in equation 3.2, which are constant, we have to minimize $\sum_\alpha (\lambda_1 \mu(\alpha)^2 + \lambda_2 \mu(\alpha)) = \lambda_1 \sum_\alpha (\mu(\alpha) + \lambda)^2 +$ const, where $\lambda \triangleq \frac{\lambda_2}{2\lambda_1}$. Choosing the weighting function $w(\alpha) = 1$ and assuming $\lambda_1 > 0$ as the experimental data indicate (λ_1 is the slope of the regression line in Fig 1B), this is equivalent to minimizing the dot product $\langle \mu + \lambda, \mu + \lambda \rangle$.

Let $\mathbf{e}_{1,...,D}$ be an orthonormal basis of $\mathbb{R}^D$, with respect to which $\bar{\mathbf{r}}$ has coordinates $\bar{\mathbf{r}}^T \mathbf{e}_1, \ldots, \bar{\mathbf{r}}^T \mathbf{e}_D$ and $\mathbf{u}(\alpha)$ has coordinates $\mathbf{u}(\alpha)^T \mathbf{e}_1, \ldots, \mathbf{u}(\alpha)^T \mathbf{e}_D$. Then we can define $r_{1,...,D} \triangleq \bar{\mathbf{r}}^T \mathbf{e}_1, \ldots, \bar{\mathbf{r}}^T \mathbf{e}_D, r_{D+1} \triangleq C, g_{1...D}(\alpha) \triangleq \mathbf{u}(\alpha)^T \mathbf{e}_1, \ldots,$

---

[10] $\lfloor x \rfloor = x$ for $x \geq 0$ and 0 otherwise. Similarly, $\lceil x \rceil = x$ for $x < 0$ and 0 otherwise.

$\mathbf{u}(\alpha)^T \mathbf{e}_D$, $g_{D+1}(\alpha) \triangleq 1$, $N \triangleq D + 1$ or D depending on whether the cocontraction command is specified.

With these definitions, the problem is in the form of equation 4.1, theorem 1 applies, and we are guaranteed that the unique optimal activation profile is $\mu^*(\alpha) = \lfloor \sum_n a_n g_n(\alpha) - \lambda \rfloor$ as long as we can find $a_{1,\ldots,N} \in \mathbb{R}$ for which μ^* satisfies all constraints.

Why is that function a cosine? The function $g_n(\alpha) = \mathbf{u}(\alpha)^T \mathbf{e}_n$ is the cosine of the angle between unit vectors $\mathbf{u}(\alpha)$ and $\mathbf{e}_n$. A linear combination of D-dimensional cosines is also a cosine: $\sum_n a_n g_n(\alpha) = \sum_n a_n \mathbf{u}(\alpha)^T \mathbf{e}_n = \mathbf{u}(\alpha)^T (\sum_n a_n \mathbf{e}_n)$, and thus $\mu^*(\alpha) = \lfloor \mathbf{u}(\alpha)^T \mathbf{E} - \lambda \rfloor$ for $\mathbf{E} = \sum_n a_n \mathbf{e}_n$. When C is specified, we have $\mu^*(\alpha) = \lfloor \mathbf{u}(\alpha)^T \mathbf{E} + a_{D+1} - \lambda \rfloor$ since $g_{D+1}(\alpha) = 1$. Note that if we are given $\mathbf{E}$, the constants a are simply $a_n = \mathbf{E}^T \mathbf{e}_n$ since the basis $\mathbf{e}_{1,\ldots,D}$ is orthonormal.

To summarize the results so far, we showed that the minimum generalized length $\langle \mu + \lambda, \mu + \lambda \rangle$ of the nonnegative activation profile $\mu(\alpha)$ subject to linear equality constraints $\langle \mu, g_{1,\ldots,N} \rangle = r_{1,\ldots,N}$ is achieved for a truncated linear combination $\lfloor \sum_n a_n g_n(\alpha) - \lambda \rfloor$ of the projection functions $g_{1,\ldots,N}$. Given a mean force $\bar{\mathbf{r}}$ and optionally a cocontraction command C, this generalized length is proportional to the variance of the net muscle force, with the projection functions being cosines. Since a linear combination of cosines is a cosine, the optimal activation profile is a truncated cosine.

In the rest of this section, we compute the optimal activation profile in two special cases. In each case, all we have to do is construct—by whatever means—a function of the specified form that satisfies all constraints. Theorem 1 then guarantees that we have found the unique global minimum.

4.2 Uniform Distribution of Force Directions in $\mathbb{R}^D$. For convenience, we will work with a continuous set of force generators $\Omega = \mathcal{S}_D$, the unit sphere embedded in $\mathbb{R}^D$. The summation signs will be replaced by integrals, and the force generator index α will be assumed to cover $\mathcal{S}_D$ uniformly, that is, the distribution of force directions is uniform. The normalization constant becomes the surface area[11] of the unit sphere $|\mathcal{S}_D| = 2\pi^{\frac{D}{2}}/\Gamma(\frac{D}{2})$. The unit vector $\mathbf{u}(\alpha) \in \mathbb{R}^D$ corresponds to point α on $\mathcal{S}_D$. The goal is to find a truncated cosine function that satisfies the constraints $|\mathcal{S}_D|^{-1} \int_{\mathcal{S}_D} \mu^*(\alpha) \mathbf{u}(\alpha) d\alpha = \bar{\mathbf{r}}$ and optionally $|\mathcal{S}_D|^{-1} \int_{\mathcal{S}_D} \mu^*(\alpha) d\alpha = C$.

We will look for a solution with axial symmetry around $\bar{\mathbf{r}}$, that is, a $\mu^*(\alpha)$, that depends only on the angle between the vectors $\bar{\mathbf{r}}$ and $\mathbf{u}(\alpha)$ rather than the actual direction $\mathbf{u}(\alpha)$. This problem can be transformed into a problem on the circle in $\mathbb{R}^2$ by correcting for the area of $\mathcal{S}_D$ being mapped into each point on the circle.

---

[11] The first few values of $|\mathcal{S}_D|$ are: $|\mathcal{S}_{1,\ldots,7}| = (2, 2\pi, 4\pi, 2\pi^2, \frac{8}{3}\pi^2, \pi^3, \frac{16}{15}\pi^3)$. Numerically $|\mathcal{S}_D|$ decreases for $D > 7$.

The set of unit vectors $\mathbf{u} \in \mathbb{R}^D$ at angle α away from a given vector $\bar{\mathbf{r}} \in \mathbb{R}^D$ is a sphere in $\mathbb{R}^{D-1}$ with radius $|\sin(\alpha)|$. Therefore, for any function $f: \mathcal{S}_D \to \mathbb{R}$ with axial symmetry around $\bar{\mathbf{r}}$, we have $\int_{\mathcal{S}_D} f = \frac{1}{2}|\mathcal{S}_{D-1}| \int_{-\pi}^{\pi} f(\alpha)|\sin(\alpha)|^{D-2} d\alpha$, where the correction factor $|\mathcal{S}_{D-1}||\sin(\alpha)|^{D-2}$ is the surface area of an $\mathbb{R}^{D-1}$ sphere with radius $|\sin(\alpha)|$.

Without loss of generality, $\bar{\mathbf{r}}$ can be assumed to point along the positive x-axis: $\bar{\mathbf{r}} = [R\ 0]$. For given dimensionality D, define the weighting function $w_D(\alpha) \triangleq \frac{1}{2}|\mathcal{S}_{D-1}||\sin(\alpha)|^{D-2}$ for $\alpha \in [-\pi; \pi]$, which as before defines the dot product $\langle u, v \rangle_D \triangleq |\mathcal{S}_D|^{-1} \int_{-\pi}^{\pi} u(\alpha)v(\alpha)w_D(\alpha)d\alpha$. The projection functions on the circle in $\mathbb{R}^2$ are $g_1(\alpha) = \cos(\alpha)$, $g_2(\alpha) = \sin(\alpha)$, and optionally $g_3(\alpha) = 1$.

Thus, μ^* has to satisfy $\langle \mu^*, \cos \rangle_D = R$, $\langle \mu^*, \sin \rangle_D = 0$, and optionally $\langle \mu^*, 1 \rangle_D = C$. Below, we set $a_2 = 0$ and find constants $a_1, a_3 \in \mathbb{R}$ for which the function $\mu^*(\alpha) = \lfloor a_1 \cos(\alpha) + a_3 \rfloor$ satisfies those constraints. Since $\langle \lfloor a_1 \cos(\alpha) + a_3 \rfloor, \sin \rangle_D = 0$ for any a_1, a_3, we are concerned only with the remaining two constraints.

Note that $\langle \mu^*, \cos \rangle_D \leq \langle \mu^*, 1 \rangle_D$ and therefore $R \leq C$ whenever C is specified. Also, from the definition of $w_D(\alpha)$ and the identity $D \int \cos^2 \sin^{D-2} = \sin^{D-1} \cos + \int \sin^{D-2}$, it follows that $\langle \cos, 1 \rangle_D = 0$, $\langle 1, 1 \rangle_D = 1$, and $\langle \cos, \cos \rangle_D = \frac{1}{D}$.

4.2.1 Specified Cocontraction C.   We are looking for a function of the form $\mu^*(\alpha) = \lfloor a_1 \cos(\alpha) + a_3 \rfloor$ that satisfies the equality constraints. For $a_3 \geq a_1$, this function is a full cosine. Using the above identities, we find that the constraints $R = \langle \mu^*, \cos \rangle_D = a_1 \langle \cos, \cos \rangle_D + a_3 \langle 1, \cos \rangle_D$ and $C = \langle \mu^*, 1 \rangle_D = a_1 \langle \cos, 1 \rangle_D + a_3 \langle 1, 1 \rangle_D$ are satisfied when $a_1 = DR$ and $a_3 = C$.

Thus, the optimal solution is a full cosine when $\frac{C}{R} \geq D$ (corresponding to $a_3 \geq a_1$). When $\frac{C}{R} < D$, a full cosine solution cannot be found; thus, we look for a truncated cosine solution. Let the truncation point be $\alpha = \pm t$, that is, $a_3 = -a_1 \cos(t)$. To satisfy all constraints, t has to be the root of the trigonometric equation,

$$\frac{C}{R} = \frac{\sin(t)^{D-1}/(D-1) - \cos(t)I_D(t)}{I_D(t) - \cos(t)\sin(t)^{D-1}/D(D-1)},$$

where $I_D(t) \triangleq \int_0^t \sin(\alpha)^{D-2} d\alpha$. That integral can be evaluated analytically for any fixed D. Once t is computed numerically, the constant a_1 is given by $a_1 = R\frac{|\mathcal{S}_D|}{|\mathcal{S}_{D-1}|}(I_D(t) - \cos(t)\sin(t)^{D-1}/D(D-1))^{-1}$. It can be verified that the above trigonometric equation has a unique solution for any value of $\frac{C}{R}$ in the interval $(1, D)$. Values smaller than 1 are inadmissible because $R \leq C$.
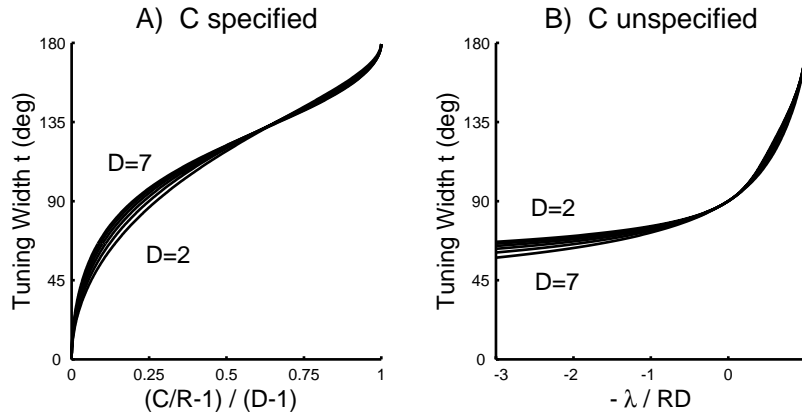
A) C specified

B) C unspecified

Figure 3: (A) The truncation point–optimal tuning width t computed from equation 4.2 for $D = 2, \ldots, 7$. Since the solution is a truncated cosine when $1 < \frac{C}{R} < D$, it is natural to scale the x-axis of the plot as $\frac{C/R-1}{D-1}$, which varies from 0 to 1 regardless of the dimensionality D. For $\frac{C}{R} \geq D$, we have the full cosine solution, which technically corresponds to $t = 180$. (B) The truncation point–optimal tuning width t computed from equation 4.3 for $D = 2, \ldots, 7$. Since the solution is a truncated cosine when $\frac{-\lambda}{R} < D$, it is natural to scale the x-axis of the plot as $\frac{-\lambda}{RD}$, which varies from $-\infty$ to 1 regardless of the dimensionality D. For $\frac{-\lambda}{R} \geq D$, we have the full cosine solution: $t = 180$.

Summarizing the solution,

$$\mu^*(\alpha) = \begin{cases} DR\cos(\alpha) + C & : \dfrac{C}{R} \geq D \\ a_1 \lfloor \cos(\alpha) - \cos(t) \rfloor & : \dfrac{C}{R} < D. \end{cases} \tag{4.2}$$

In Figure 3A we have plotted the optimal tuning width t in different dimensions, for the truncated cosine case $\frac{C}{R} < D$.

4.2.2 Unspecified Cocontraction C.

In this case, $a_3 = -\lambda$, that is, $\mu^*(\alpha) = \lfloor a_1 \cos(\alpha) - \lambda \rfloor$. For $-\lambda \geq a_1$, the solution is a full cosine, and $a_1 = DR$ as before. When $-\lambda < DR$, the solution is a truncated cosine. Let the truncation point be $\alpha = \pm t$. Then $a_1 = \frac{\lambda}{\cos(t)}$, and t has to be the root of the trigonometric equation:

$$\frac{\lambda}{\cos(t)} = R \frac{|\mathcal{S}_D|}{|\mathcal{S}_{D-1}|} (I_D(t) - \cos(t)\sin(t)^{D-1}/D(D-1))^{-1}.$$

Note that a_1 as a function of t is identical to the previous case when C was fixed, while the equation for t is different.

Summarizing the solution:

$$\mu^*(\alpha) = \begin{cases} DR\cos(\alpha) - \lambda & : \dfrac{-\lambda}{R} \geq D \\[2mm] a_1\lfloor\cos(\alpha) - \cos(t)\rfloor & : \dfrac{-\lambda}{R} < D \end{cases}. \tag{4.3}$$

In Figure 3B we have plotted the optimal tuning width t in different dimensions, for the truncated cosine case $\frac{-\lambda}{R} < D$.

Comparing the curves in Figure 3A and Figure 3B, we notice that in both cases, the optimal tuning width is rather large (it is advantageous to activate multiple force generators), except in Figure 3A for $C/R \approx 1$. From the triangle inequality, a solution $\mu(\alpha)$ exists only when $C \geq R$, and $C = R$ implies $\mu(\alpha \neq 0) = 0$. Thus, a small C "forces" the activation profile to become a delta function. But as soon as that constraint is relaxed, the width of the optimal solution increases sharply. Note also that in both figures, the dimensionality D makes little difference after appropriate scaling of the abscissa.

4.3 Arbitrary Distribution of Force Directions in $\mathbb{R}^2$. For a uniform distribution of force directions, it was possible to replace the term $\sum_\alpha (\mu(\alpha)+\lambda)^2$ with $\int_{\mathcal{S}_D}(\mu(\alpha) + \lambda)^2 d\alpha$ in equations 3.2. If the distribution is not uniform but instead is given by some density function $w(\alpha)$, we have to take that function into account and find the activation profile μ^* that minimizes $|\mathcal{S}_D|^{-1}\int_{\mathcal{S}_D}(\mu^*(\alpha) + \lambda)^2 w(\alpha)d\alpha$ subject to $|\mathcal{S}_D|^{-1}\int_{\mathcal{S}_D}\mu^*(\alpha)\mathbf{u}(\alpha)w(\alpha)d\alpha = \bar{\mathbf{r}}$ and optionally $|\mathcal{S}_D|^{-1}\int_{\mathcal{S}_D}\mu^*(\alpha)w(\alpha)d\alpha = C$. Theorem 1 still guarantees that the optimal μ^* is a truncated cosine, assuming we can find a truncated cosine satisfying the constraints. It is not clear how to do that for arbitrary dimensionality D and arbitrary density w, so we address only the case $D = 2$.

For arbitrary $w(\alpha)$ and $D = 2$, the solution is in the form $\mu^*(\alpha) = \lfloor a_1\cos(\alpha) + a_2\sin(\alpha) + a_3\rfloor$. When C is not specified, we have $a_3 = -\lambda$. Here we evaluate these parameters only when C is specified and large enough to ensure a full cosine solution. The remaining cases can be handled using techniques similar to the previous sections. Expanding $w(\alpha)$ in a Fourier series, $w(\alpha) = \frac{u_0}{2} + \sum_{n=1}^{\infty}(u_n\cos n\alpha + v_n\sin n\alpha)$ and solving the system of linear equations given by the constraints, we obtain

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \dfrac{u_0 + u_2}{2} & \dfrac{v_2}{2} & u_1 \\ \dfrac{v_2}{2} & \dfrac{u_0 - u_2}{2} & v_1 \\ u_1 & v_1 & u_0 \end{bmatrix}^{-1} \begin{bmatrix} 2R \\ 0 \\ 2C \end{bmatrix}.$$

The optimal μ^* depends only on the Fourier coefficients of $w(\alpha)$ up to order 2; the higher-order terms do not affect the minimization problem. In

the previous sections, $w(\alpha)$ was equal to the dimensionality correction factor $|\sin(\alpha)|^{D-2}$, in which case only u_0 and u_2 were nonzero, the above matrix became diagonal, and thus we had $a_1 \sim R$, $a_2 = 0$, $a_3 \sim C$.

Note that $\mu^*(\alpha)$ is the optimal activation profile over the set of force generators for fixed mean force. In the case of uniformly distributed force directions, this also described the tuning function of an individual force generator for varying mean force direction, since $\mu^*(\alpha)$ was centered at 0 and had the same shape regardless of force direction. That is no longer true here. Since $w(\alpha)$ can be asymmetric, the directions of the force generator and the mean force matter (as illustrated in Figures 4a and 4b). The tuning functions of several force generators at different angles from the peak of $w(\alpha)$ are plotted in Figures 4a and 4b. The direction of maximal activation rotates away from the generator force direction and toward the short axis of $w(\alpha)$, for generators whose force direction lies in between the short and long axes of $w(\alpha)$. This effect has been observed experimentally for planar arm movements, where the distribution of muscle lines of action is elongated along the hand-shoulder axis (Cisek & Scott, 1998). In that case, muscles are maximally active when the net force is rotated away from their mechanical line of action, toward the short axis of the distribution. The same effect is seen in wrist muscles, where the distribution of lines of action is again asymmetric (Hoffman & Strick, 1999).

The tuning modulation (difference between the maximum and minimum of the tuning curve) also varies systematically, as shown in Figures 4a and 4b. Such effects are more difficult to detect experimentally, since that would require comparisons of the absolute values of signals recorded from different muscles or neurons.

5 Some Extensions

5.1 Optimal Force Bias. The optimal force bias can be found by solving equation 3.1: minimize $V^*(\bar{\mathbf{r}}) + B(\bar{\mathbf{r}})$ w.r.t. $\bar{\mathbf{r}}$. We will solve it analytically only for a uniform distribution of force directions in $\mathbb{R}^D$ and when the minimum in equation 3.2 is a full cosine. It can be shown using equations 4.1 and 4.2 that for both C, specified and unspecified, the variance term dependent on μ^* in equation 3.2 is $\frac{\lambda_1 D}{|\mathcal{S}_D|} R^2$. It is clear that the optimal mean force $\bar{\mathbf{r}}$ is parallel to the desired force $\mathbf{f}$, and all we have to find is its magnitude $R = \|\bar{\mathbf{r}}\|$. Then to solve equation 3.1, we have to minimize w.r.t. R the following expression: $\frac{\lambda_1 D}{|\mathcal{S}_D|} R^2 + \lambda_1 \lambda_3 R^2 + (R - \|\mathbf{f}\|)^2$. Setting the derivative to 0, the minimum is achieved for

$$\|\bar{\mathbf{r}}\| = \frac{\|\mathbf{f}\|}{1 + \lambda_1 \lambda_3 + \lambda_1 D / |\mathcal{S}_D|}.$$

Thus, for positive λ_1, λ_3 the optimal mean force magnitude $\|\bar{\mathbf{r}}\|$ is smaller than the desired force magnitude $\|\mathbf{f}\|$, and the optimal bias $\|\bar{\mathbf{r}}\| - \|\mathbf{f}\|$ increases linearly with $\|\mathbf{f}\|$.

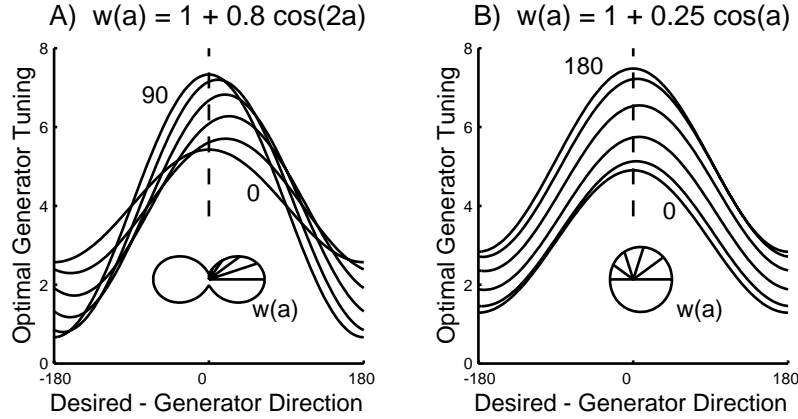A) w(a) = 1 + 0.8 cos(2a)   B) w(a) = 1 + 0.25 cos(a)

Figure 4: Optimal tuning of generators whose force directions point at different angles relative to the peak of the distribution $w(a)$. 0 corresponds to the peak of $w(a)$, and 90 (180, respectively) corresponds to the minimum of $w(a)$. The polar plots show $w(a)$, and the lines inside indicate the generator directions plotted in each figure. We used $R = 1$, $C = 5$. (A) $w(a)$ has a second-order harmonic. In this case, the direction of maximal activation for generators near 45 rotates toward the short axis of $w(a)$. The optimal tuning modulation increases for generators near 90. (B) $w(a)$ has a first-order harmonic. In this case, the rotation is smaller, and the tuning curves near the short axis of $w(a)$ shift upward rather than increasing their modulation.

5.2 Error-Effort Trade-off. Our formulation of the optimal control problem facing the motor system assumed that the only quantity being minimized is error (see equation 3.1). It may be more sensible, however, to minimize a weighted sum of error and effort, because avoiding fatigue in the current task can lead to smaller errors in tasks performed in the future. Indeed, we have found evidence for error+effort minimization in movement tasks (Todorov, 2001). To allow this possibility here, we consider a modified cost function of the form

$$\mathrm{E}_{z|\mu}[(\mathbf{r} - \mathbf{f})^T(\mathbf{r} - \mathbf{f})] + \beta |\Omega|^{-2} \sum_{\alpha \in \Omega} \mu(\alpha)^2.$$

The only change resulting from the inclusion of the activation penalty term is that the variance V previously given by equation 3.2 now becomes

$$V = |\Omega|^{-2} \sum_{\alpha \in \Omega} ((\lambda_1 + \beta)\mu(\alpha)^2 + \lambda_2 \mu(\alpha)) + \lambda_1 \lambda_3 \bar{\mathbf{r}}^T \bar{\mathbf{r}} + \lambda_2 \lambda_3 \bar{\mathbf{r}}^T \mathbf{U}.$$

Thus, the results in section 4 remain unaffected (apart from the substitution $\lambda_1 \leftarrow \lambda_1 + \beta$), and the optimal tuning curve is the same as before. The only

effect of the activation penalty is to increase the force bias. The optimal $\|\bar{\mathbf{r}}\|$ computed in section 5.1 now becomes

$$\|\bar{\mathbf{r}}\| = \frac{\|\mathbf{f}\|}{1 + \lambda_1\lambda_3 + (\lambda_1 + \beta)D/|\mathcal{S}_D|}.$$

Thus, the optimal bias $\|\bar{\mathbf{r}}\| - \|\mathbf{f}\|$ increases with the weight β of the activation penalty. This can explain why the experimentally observed bias in Figure 1A was larger than predicted by minimizing error alone.

5.3 Nonhomogeneous Noise Correlations. Thus far, we allowed only homogeneous correlations (λ_3) among noise terms affecting different generators. Here, we consider an additional correlation term (λ_4) that varies with the angle between two generators. The noise covariance model $\mathrm{Cov}_{z|\mu}[z(\alpha), z(\beta)]$ now becomes

$$\left(\lambda_1\mu(\alpha)\mu(\beta) + \lambda_2\frac{\mu(\alpha) + \mu(\beta)}{2}\right)(\delta_\alpha^\beta + \lambda_3 + 2\lambda_4\mathbf{u}(\beta)^T\mathbf{u}(\alpha)).$$

We focus on the case when the force generators are uniformly distributed in a two-dimensional work space ($D = 2$), the mean force is $\bar{\mathbf{r}} = [R\ 0]$ as before, the cocontraction level C is specified, and $\frac{C}{R} \geq D$. Using the identities $\mathbf{u}(\beta)^T\mathbf{u}(\alpha) = \cos(\alpha - \beta)$ and $2\cos^2(\alpha - \beta) = 1 + \cos(2\alpha - 2\beta)$, the force variance V previously given by equation 3.2 now becomes

$$\frac{1}{4\pi^2}\int(\lambda_1\mu(\alpha)^2 + \lambda_2\mu(\alpha))d\alpha + \lambda_1\lambda_3\bar{\mathbf{r}}^T\bar{\mathbf{r}} + \frac{\lambda_1\lambda_4}{4}(p_2^2 + q_2^2) + \lambda_1\lambda_4C^2 + \lambda_2\lambda_4C,$$

where $p_2 = \frac{1}{\pi}\int\mu(\alpha)\cos(2\alpha)d\alpha$ and $q_2 = \frac{1}{\pi}\int\mu(\alpha)\sin(2\alpha)d\alpha$ are the second order coefficients in the Fourier series $\mu(\alpha) = \frac{p_0}{2} + \sum_{n=1}^{\infty}(p_n\cos n\alpha + q_n\sin n\alpha)$.

The integral term in V can be expressed as a function of the Fourier coefficients using Parseval's theorem. The constraints on $\mu(\alpha)$ are $\frac{1}{2\pi}\int\mu(\alpha)d\alpha = C$, $\frac{1}{2\pi}\int\mu(\alpha)\cos(\alpha)d\alpha = R$, and $\frac{1}{2\pi}\int\mu(\alpha)\sin(\alpha)d\alpha = 0$. These constraints specify the p_0, p_1, and q_1 Fourier coefficients. Collecting all unconstrained terms in V yields

$$V = \frac{\lambda_1}{4}\left(\lambda_4 + \frac{1}{\pi}\right)(p_2^2 + q_2^2) + \frac{\lambda_1}{4\pi}\sum_{n=3}^{\infty}(p_n^2 + q_n^2) + \mathrm{const}(C, \bar{\mathbf{r}}).$$

Since the parameter λ_1 corresponding to the slope of the regression line in Figure 1B is positive, the above expression is a sum of squares with positive weights when $\lambda_4 > -\frac{1}{\pi}$. The unique minimum is then achieved when $p_{2,\dots,\infty} = q_{2,\dots,\infty} = 0$, and therefore the optimal tuning curve is $\mu(\alpha) = DR\cos(\alpha) + C$ as before.

If nonhomogeneous correlations are present, one would expect muscles pulling in similar directions to be positively correlated ($\lambda_4 > 0$), as simultaneous EMG recordings indicate (Stephens, Harrison, Mayston, Carr, & Gibbs, 1999). This justifies the assumption $\lambda_4 > -\frac{1}{\pi}$.

5.4 Alternative Cost Functions. We assumed that the cost being minimized by the motor system is the square error of the force output. While a square error cost is common to most optimization models in motor control (see section 6), it is used for analytical convenience without any empirical support. This is not a problem for phenomenological models that simply look for a quantity whose minimum happens to match the observed behavior. But if we are to construct more principled models and claim some correspondence to a real optimization process in the motor system, it is necessary to confirm the behavioral relevance of the chosen cost function. How can we proceed in the absence of such empirical confirmation? Our approach is to study alternative cost functions, obtain model predictions through numerical simulation, and show that the particular cost function being chosen makes little difference.

Throughout this section, we assume that C is specified, the work space is two-dimensional, and the target force (without loss of generality) is $\mathbf{f} = [R\,0]$. The cost function is now

$$\text{Cost}_p(\mu) = \text{E}(\|\mathbf{r} - \mathbf{f}\|^p).$$

We find numerically the optimal activations $\mu_{1,\dots,15}$ for 15 uniformly distributed force generators. The noise terms $z_{1,\dots,15}$ are assumed independent, with probability distribution matching the experimental data. In order to generate such noise terms, we combined the data for each instructed force level (all subjects, one-hand condition), subtracted the mean, divided by the standard deviation, and pooled the data from all force levels. Samples from the distribution of z_i were then obtained as $z_i = \lambda_1 \mu_i s$, where s was sampled with replacement from the pooled data set. The scaling constant was set to $\lambda_1 = 0.2$. It could not be easily estimated from the data (because subjects used multiple muscles), but varying it from 0.2 to 0.1 did not affect the results presented here, as expected from section 4.

To find the optimal activations, we initialized $\mu_{1,\dots,15}$ randomly, and minimized the Monte Carlo estimate of $\text{Cost}_p(\mu)$ using BFGS gradient-descent with numerically computed gradients (fminunc in the Maltab Optimization Toolbox). The constraints $\mu_i \geq 0$ and $\frac{1}{15}\sum \mu_i = C$ were enforced by scaling and using the absolute value of μ_i inside the estimation function. A small cost proportional to $|\frac{1}{15}\sum \mu_i - C|$ was added to resolve the scaling ambiguity. To speed up convergence, a fixed $15 \times 100{,}000$ random sample from the experimental data was used in each minimization run.

The average of the optimal tuning curves found in 40 runs of the algorithm (using different starting points and random samples) is plotted in
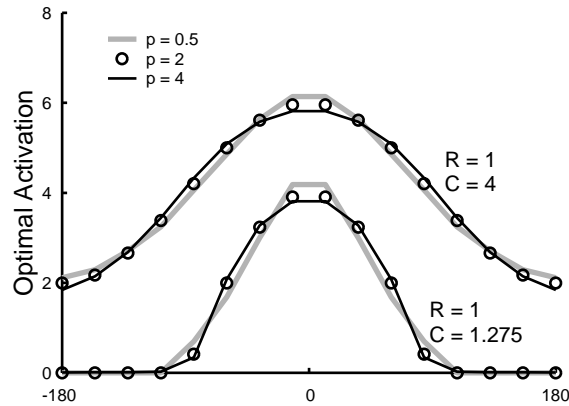
Figure 5: The average of 40 optimal tuning curves, for $p = 0.2$ and $p = 4$. The different tuning curves found in multiple runs were similar. The solution for $p = 2$ was computed using the results in section 4.

Figure 5, for $p = 0.5$ and $p = 4$. The optimal tuning curve with respect to the quadratic cost ($p = 2$) is also shown. For both full and truncated cosine solutions, the choice of cost function made little difference. We have repeated this analysis with gaussian noise and obtained very similar results.

It is in principle possible to compare the different curves in Figure 5 to experimental data and try to identify the true cost function used by the motor system. However, the differences are rather small compared to the noise in empirically observed tuning curves, so this analysis is unlikely to produce unambiguous results.

5.5 Movement Velocity and Displacement Tuning. The above analysis explains cosine tuning with respect to isometric force. To extend our results to dynamic conditions and address movement velocity and displacement tuning, we have to take into account the fact that muscle force production is state dependent. For a constant level of activation, the force produced by a muscle varies with its length and rate of change of length (Zajac, 1989), decreasing in the direction of shortening. The only way the CNS can generate a desired net muscle force during movement is to compensate for this dependence: since muscles pulling in the direction of movement are shortening, their force output for fixed neural input drops, and so their neural input has to increase. Thus, muscle activation has to correlate with movement velocity and displacement (Todorov, 2000).

Now consider a short time interval in which neural activity can change, but all lengths, velocities, and forces remain roughly constant. In this setting, the analysis from the preceding sections applies, and the optimal tun-

ing curve with respect to movement velocity and displacement is again a
(truncated) cosine. While the relationship between muscle force and activa-
tion can be different in each time interval, the minimization problem itself
remains the same; thus, each solution belongs to the family of truncated
cosines described above. The net muscle force that the CNS attempts to
generate in each time interval can be a complex function of the estimated
state of the limb and the task goals. This complexity, however, does not
affect our argument: we are not asking how the desired net muscle force is
computed but how it can be generated accurately once it has been computed.

The quasi-static setting considered here is an approximation, which is
justified because the neural input is low-pass-filtered before generating
force (the relationship between EMG and muscle force is well modeled by a
second-order linear filter with time constants around 40 msec; Winter, 1990),
and lengths and velocities are integrals of the forces acting on the limb, so
they vary even more slowly compared to the neural input. Replacing this
approximation with a more detailed model of optimal movement control is
a topic for future work.

6 Discussion

In summary, we developed a model of noisy force production where opti-
mal tuning is defined in terms of expected net force error. We proved that
the optimal tuning curve is a (possibly truncated) cosine, for a uniform dis-
tribution $w(\alpha)$ of force directions in $\mathbb{R}^D$ and for an arbitrary distribution
$w(\alpha)$ of force directions in $\mathbb{R}^2$. When both $w(\alpha)$ and D are arbitrary, the
optimal tuning curve is still a truncated cosine, provided that a truncated
cosine satisfying all constraints exists. Although the analytical results were
obtained under the assumptions of quadratic cost and homogeneously cor-
related noise, it was possible to relax these assumptions in special cases.
Redefining optimal tuning in terms of error+effort minimization did not
affect our conclusions.

The model makes three novel and somewhat surprising predictions. First,
the model predicts a relationship between the shape of the tuning curve $\mu(\alpha)$
and the cocontraction level C. According to equation 4.2, when C is large
enough, the optimal tuning curve $\mu(\alpha) = DR\cos(\alpha) + C$ is a full cosine,
which scales with the magnitude of the net force R and shifts with C. But
when C is below the threshold value DR, the optimal tuning curve is a
truncated cosine, which becomes sharper as C decreases. Thus, we would
expect to see sharper-than-cosine tuning curves in the literature. Such ex-
amples can indeed be found in Turner et al. (1995) and Hoffman and Strick
(1999). A more systematic investigation in M1 (Amirikian & Georgopoulos,
2000) revealed that the tuning curves of most cells were better fit by sharper-
than-cosine functions, presumably because of the low cocontraction level.
We recently tested the above prediction using both M1 and EMG data and
found that cells and muscles that appear to have higher contributions to

the cocontraction level also have broader tuning curves, whose average is indistinguishable from a cosine (Todorov et al., 2000). This prediction can be tested more directly by asking subjects to generate specified net forces and simultaneously achieve different cocontraction levels.

Second, under nonuniform distributions of force directions, the model predicts a misalignment between preferred and force directions, while the tuning curves remain cosine. This effect has been observed by Cisek and Scott (1998) and Hoffman and Strick (1999). Note that a nonuniform distribution of force directions does not necessitate misalignment; instead, the asymmetry can be compensated by using skewed tuning curves.

Third, our analysis shows that optimal force production is negatively biased; the bias is larger when fewer force generators are active and increases with mean force. The measured bias was larger than predicted from error minimization alone, which suggests that the motor system minimizes a combination of error and effort in agreement with results we have recently obtained in movement tasks (Todorov, 2001).

The model for the first time demonstrates how cosine tuning could result from optimizing a meaningful objective function: accurate force production. Another model proposed recently (Zhang & Sejnowski, 1999a) takes a very different approach. It assumes a universal rule for encoding motion information in both sensory and motor areas,[12] which gives rise to cosine tuning. Its main advantage is that tuning for movement direction can be treated in the same framework in all parts of the nervous system, regardless of whether the motion signal is related to a body part or an external object perceived visually. But that model has two disadvantages: (1) it cannot explain cosine tuning with direction of force and displacement in the motor system, and (2) cosine tuning is explained with a new encoding rule that remains to be verified experimentally. If the new encoding rule is confirmed, it would provide a mechanistic explanation of cosine tuning that does not address the question of optimality. In that sense, the model of Zhang and Sejnowski (1999a) can be seen as being complementary to ours.

6.1 Origins of Neuromotor Noise. The origin and scaling properties of neuromotor noise are of central importance in stochastic optimization models of the motor system. The scaling law relating the mean and standard deviation of the net force was derived experimentally. What can we say about the neural mechanisms responsible for this type of noise? Very little, unfortunately.

Although a number of studies on motor tremor have analyzed the peaks in the power spectrum and how they are affected by different experimen-

---

[12] Assume each cell has a "hidden" function $\Phi(x)$ and encodes movement in $x \in \mathbb{R}^D$ by firing in proportion to $d\Phi(x(t))/dt$. From the chain rule $d\Phi/dt = \partial\Phi/\partial x \, . \, dx/dt = \nabla\Phi \, . \, \dot{x}$. This is the dot product of a cell-specific "preferred direction" $\nabla\Phi$ and the movement velocity vector $\dot{x}$—thus, cosine tuning for movement velocity.

tal manipulations, no widely accepted view of their origin has emerged (McAuley et al., 1997). Possible explanations include noise in the central drive, oscillations arising in spinal circuits, effects of afferent input, and mechanical resonance.

One might expect the noise in the force output to reflect directly the noise in the descending M1 signals, in agreement with the finding that magnetoencephalogram fluctuations recorded over M1 are synchronous with EMG activity in contralateral muscles (Conway et al., 1995). On the level of single cells and muscles, however, this relationship is quite complicated. Cells in M1 (and most other areas of cortex) are well modeled as Poisson processes with coefficients of variation (CV) around 1 (Lee, Port, Kruse, & Georgopoulos, 1998). For a Poisson process, the spike count in a fixed interval has variance (rather than standard deviation) linear in the mean. The firing patterns of motoneurons are nothing like Poisson processes. Instead, motoneurons fire much more regularly, with CVs around 0.1 to 0.2 (DeLuca, 1995). Furthermore, muscle force is controlled to a large extent by recruiting new motor units, so noise in the force output may arise from the motor unit recruitment mechanisms, which are not very well understood. Other physiological mechanisms likely to affect the output noise distribution are recurrent feedback through Renshaw cells (which may serve as a decorrelating mechanism; Maltenfort, Heckman, & Rymer, 1998), as well as plateau potentials (caused by voltage-activated calcium channels) that may cause sustained firing of motoneurons in the absence of synaptic input (Kiehn & Eken, 1997). Also, muscle force is not just a function of motoneuronal firing rate, but depends significantly on the sequence of interspike intervals (Burke, Rudomin, & Zajac, 1976).[13] Thus, although the mean firing rates of M1 cells seem to contribute additively to the mean activations of muscle groups (Todorov, 2000), the small timescale fluctuations in M1 and muscles have a more complex relationship.

The motor tremor illustrated in Figure 1B should not be thought of as being the only source of noise. Under dynamic conditions, various calibration errors (such as inaccurate internal estimates of muscle fatigue, potentiation, length, and velocity dependence) can have a compound effect resembling multiplicative noise. This may be why the errors observed in dynamic force tasks (Schmidt et al., 1979) as well as reaching without vision (Gordon, Ghilardi, Cooper, & Ghez, 1994) are substantially larger than what the slopes in Figure 1B would predict.

6.2 From Muscle Tuning to M1 Cell Tuning. Since M1 cells are synaptically close to motoneurons (in some cases, the projection can even be monosynaptic; Fetz & Cheney, 1980), their activity would be expected to

---

[13] Because of this nonlinear dependence, muscle force would be much noisier if motoneurons had Poisson firing rates, which may be why they fire so regularly.

reflect properties of the motor periphery. The defining feature of a muscle is its line of action (determined by the tendon insertion points), in the same way that the defining feature of a photoreceptor is its location on the retina. A fixed line of action implies a preferred direction, just like a fixed retinal location implies a spatially localized receptive field. Thus, given the properties of the periphery, the existence of preferred directions in M1 is no more surprising than the existence of spatially localized receptive fields in V1.[14] Of course, directional tuning of muscles does not necessitate similar tuning in M1, in the same way that cells in V1 do not have to display spatial tuning; one can imagine, for example, a spatial Fourier transform in the retina or lateral geniculate nucleus that completely abolishes the spatial tuning arising from photoreceptors. But perhaps the nervous system avoids such drastic changes in representation, and tuning properties that arise (for whatever reason) in one area "propagate" to other densely connected areas, regardless of the direction of connectivity.

Using this line of reasoning and the fact that muscle activity has to correlate with movement velocity and displacement in order to compensate for muscle visco-elasticity (see section 5.5), we have previously explained a number of seemingly contradictory phenomena in M1 without the need to evoke abstract encoding principles (Todorov, 2000). This article adds cosine tuning to that list of phenomena. We showed here that because of the multiplicative nature of motor noise, the optimal muscle tuning curve is a cosine. This makes cosine tuning a natural choice for motor areas that are close to the motor periphery. Motor areas that are further removed from motoneurons have less of a reason to display cosine tuning. Cerebellar Purkinje cells, for example, are often tuned for a limited range of movement speeds, and their tuning curves can be bimodal (Coltz, Johnson, & Ebner, 1999).

6.3 Optimization Models in Motor Control. A number of earlier optimization models explain aspects of motor behavior as emerging from the minimization of some cost functional. The speed-accuracy trade-off known as Fitt's law has been modeled in this way (Meyer, Abrams, Kornblum, Wright, & Smith, 1988; Hoff, 1992; Harris & Wolpert, 1998). The reaching movement trajectory that minimizes expected end-point error is computed under a variety of assumptions about the control system (intermittent versus continuous, open loop versus closed loop) and the noise scaling properties (velocity- versus neural-input-dependent). While each model has advantages and disadvantages in fitting existing data, they all capture the basic logarithmic relationship between target width and movement duration.

---

[14] From this point of view, orientation tuning in V1 is surprising because it does not arise directly from peripheral properties. An equally surprising and robust phenomenon in M1 has not yet been found.

This robustness with respect to model assumptions suggests that Fitt's law indeed emerges from error minimization.

Another set of experimental results that optimization models have addressed are kinematic regularities observed in hand movements (Morasso, 1981; Lacquaniti, Terzuolo, & Viviani, 1983). While a number of physically relevant cost functions (e.g., minimum time, energy, force, impulse) were investigated (Nelson, 1983), better reconstruction of the bell-shaped speed profiles of reaching movements was obtained (Hogan, 1984) by minimizing squared jerk (derivative of acceleration). Recently, the most accurate reconstructions of complex movement trajectories were also obtained by minimizing under different assumptions the derivative of acceleration (Todorov & Jordan, 1998) or torque (Nakano et al., 1999). While these fits to experimental data are rather satisfying, the seemingly arbitrary quantity being minimized is less so.

The stochastic optimization model of Harris and Wolpert (1998) takes a more principled approach: it minimizes expected end-point error assuming that the standard deviation of neuromotor noise is proportional to the mean neural activation. Shouldn't that result in minimizing force and acceleration, which, as Nelson (1983) showed, yields unrealistic trajectories? It should, if muscle activation and force were identical, but they are not; instead muscle force is a low-pass-filtered version of activation (Winter, 1990). As a result, the neural signal under dynamic conditions contains terms related to the derivative of force, and so the model of Harris and Wolpert (1998) effectively minimizes a cost that includes jerk or torque change along with other terms. It will be interesting to find tasks where maximizing accuracy and maximizing smoothness make different predictions and test which prediction is closer to observed trajectories.

The noise model used by Harris and Wolpert (1998) is identical to ours under isometric conditions. During movement, it is not known whether noise magnitude is better fit by mean force (as in the present model) or muscle activation (as in Harris & Wolpert, 1998). Our conclusions should not be sensitive to such differences, since we do not rely on muscle low-pass filtering to explain cosine tuning. Nevertheless, it is important to establish experimentally the properties of neuromotor noise during movement.

Appendix _____

The crucial fact underlying the proof of theorem 1 is that the linear span $\mathcal{L}$ of the functions $g_{1,\ldots,N}$ is orthogonal to the hyperplane $\mathcal{P}$ defined by the equality constraints in equation 4.1.

Lemma 1. *For any $a_{1,\ldots,N} \in \mathbb{R}$ and $u, v \in \mathcal{R}(\Omega)$ satisfying $\langle u, g_{1,\ldots,N} \rangle = \langle v, g_{1,\ldots,N} \rangle = r_{1,\ldots,N}$, the $\mathcal{R}(\Omega)$ function $l(\alpha) = \sum_n a_n g_n(\alpha)$ is orthogonal to $u - v$, that is, $\langle u - v, l \rangle = 0$.*
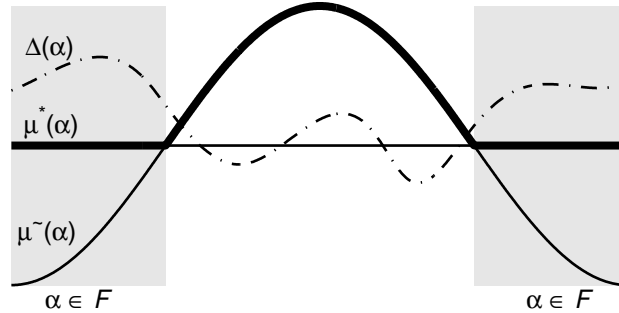
Figure 6: Illustration of the functions $\mu^*(\alpha)$, $\widetilde{\mu}(\alpha)$, $\Delta(\alpha)$ in theorem 1, case 2, with $\sum a_n g_n(\alpha) = \cos(\alpha)$. The shaded region is the set F where $\cos(\alpha) < 0$. The key point is that $\Delta(\alpha)\widetilde{\mu}(\alpha) \leq 0$ for all α.

Proof. $\langle u - v, l \rangle = \langle u, \sum_n a_n g_n \rangle - \langle v, \sum_n a_n g_n \rangle = \sum_n a_n(\langle u, g_n \rangle - \langle v, g_n \rangle) = \sum_n a_n(r_n - r_n) = 0$.

The quantity $\langle \mu + \lambda, \mu + \lambda \rangle$ we want to minimize is a generalized length, the solution μ is constrained to the hyperplane $\mathcal{P}$ orthogonal to $\mathcal{L}$, and $\mathcal{L}$ contains the origin 0. Thus, we would intuitively expect the optimal solution μ^* to be close to the intersection of $\mathcal{P}$ and $\mathcal{L}$, that is, to resemble a linear combination of $g_{1,...,N}$. The nonnegativity constraint on μ introduces complications that are handled in case 2 (see Figure 6). The proof of theorem 1 is the following:

Proof of Theorem 1. Let $\mu = \mu^* + \Delta$ for some $\Delta \in \mathcal{R}(\Omega)$ be another function satisfying all constraints in equation 4.1. Using the linearity and symmetry of the dot product, $\langle \mu + \lambda, \mu + \lambda \rangle = \langle \mu^* + \lambda, \mu^* + \lambda \rangle + 2\langle \mu^* + \lambda, \Delta \rangle + \langle \Delta, \Delta \rangle$. The term $\langle \Delta, \Delta \rangle$ is always nonnegative and becomes 0 only when $\Delta(\alpha) = 0$ for all α. Thus, to prove that μ^* is the unique optimal solution, it is sufficient to show that $\langle \mu^* + \lambda, \Delta \rangle \geq 0$. We have to distinguish two cases, depending on whether the term in the truncation brackets is positive for all α:

Case 1. Suppose $\sum_n a_n g_n(\alpha) \geq \lambda$ for all $\alpha \in \Omega$, that is, $\mu^*(\alpha) = \sum_n a_n g_n(\alpha) - \lambda$. Then $\langle \mu^* + \lambda, \Delta \rangle = \langle \sum_n a_n g_n, \mu - \mu^* \rangle = 0$ from the lemma 1.

Case 2. Consider the function $\widetilde{\mu}(\alpha) = \lceil \sum_n a_n g_n(\alpha) - \lambda \rceil$, which has the property that $\widetilde{\mu} + \mu^* = \sum_n a_n g_n - \lambda$. With this definition and using lemma 1, $0 = \langle \sum_n a_n g_n, \mu - \mu^* \rangle = \langle \widetilde{\mu} + \mu^* + \lambda, \Delta \rangle = \langle \widetilde{\mu}, \Delta \rangle + \langle \mu^* + \lambda, \Delta \rangle$. Then $\langle \mu^* + \lambda, \Delta \rangle = -\langle \widetilde{\mu}, \Delta \rangle$, and it is sufficient to show that $\langle \widetilde{\mu}, \Delta \rangle \leq 0$. Let $F \subset \Omega$ be the subset of Ω on which $\sum_n a_n g_n(\alpha) < \lambda$. Then $\widetilde{\mu}(\alpha \in F) < 0$ and $\widetilde{\mu}(\alpha \notin F) = 0$. Since $\mu = \mu^* + \Delta$ satisfies $\mu \geq 0$ and by definition

$\mu^*(\alpha \in F) = 0$, we have $\Delta(\alpha \in F) \geq 0$. The dot product $\langle \widetilde{\mu}, \Delta \rangle$ can be evaluated by parts on the two sets $\alpha \in F$ and $\alpha \notin F$. Since $\widetilde{\mu}(\alpha)\Delta(\alpha) \leq 0$ for $\alpha \in F$, and $\widetilde{\mu}(\alpha)\Delta(\alpha) = 0$ for $\alpha \notin F$, it follows that $\langle \widetilde{\mu}, \Delta \rangle \leq 0$.

Acknowledgments

References

Amirikian, B., & Georgopoulos, A. (2000). Directional tuning profiles of motor cortical cells. *Neuroscience Research*, *36*, 73–79.

Burke, R. E., Rudomin, P., & Zajac, F. E. (1976). The effect of activation history on tension production by individual muscle units. *Brain Research*, *109*, 515–529.

Caminiti, R., Johnson, P., Galli, C., Ferraina, S., & Burnod, Y. (1991). Making arm movements within different parts of space: The premotor and motor cortical representation of a coordinate system for reaching to visual targets. *Journal of Neuroscience*, *11*(5), 1182–1197.

Cisek, P., & Scott, S. H. (1998). Cooperative action of mono- and bi-articular arm muscles during multi-joint posture and movement tasks in monkeys. *Society for Neuroscience Abstracts*, *164.4.*

Clancy, E., & Hogan, N. (1999). Probability density of the surface electromyogram and its relation to amplitude detectors. *IEEE Transactions on Biomedical Engineering*, *46*(6), 730–739.

Coltz, J., Johnson, M., & Ebner, T. (1999). Cerebellar Purkinje cell simple spike discharge encodes movement velocity in primates during visuomotor tracking. *Journal of Neuroscience*, *19*(5), 1782–1803.

Conway, B. A., Halliday, D. M., Farmer, S. F., Shahani, U., Maas, P., Weir, A. I., & Rosenberg, J. R. (1995). Synchronization between motor cortex and spinal motoneuronal pool during the performance of a maintained motor task in man. *J. Physiol. (Lond.)*, *489*, 917–924.

DeLuca, C. J. (1995). Decomposition of the EMG signal into constituent motor unit action potentials. *Muscle and Nerve*, *18*, 1492–1493.

Fetz, E. E., & Cheney, P. D. (1980). Postspike facilitation of forelimb muscle activity by primate corticomotoneuronal cells. *Journal of Neurophysiology, 44*, 751–772.

Georgopoulos, A., Kalaska, J., Caminiti, R., & Massey, J. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *Journal of Neuroscience*, *2*(11), 1527–1537.

Gordon, J., Ghilardi, M. F., Cooper, S., & Ghez, C. (1994). Accuracy of planar reaching movements. *Exp. Brain Res.*, *99*, 97–130.

Harris, C. M., & Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, *394*, 780–784.

Herrmann, U., & Flanders, M. (1998). Directional tuning of single motor units. *Journal of Neuroscience*, *18*(20), 8402–8416.

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing* (pp. 77–109). Cambridge, MA: MIT Press.

Hoff, B. (1992). *A computational description of the organization of human reaching and prehension*. Unpublished doctoral dissertation, University of Southern California.

Hoffman, D. S., & Strick, P. L. (1999). Step-tracking movements of the wrist. IV. Muscle activity associated with movements in different directions. *Journal of Neurophysiology*, *81*, 319–333.

Hogan, N. (1984). An organizing principle for a class of voluntary movements. *Journal of Neuroscience*, *4*(11), 2745–2754.

Kalaska, J. F., Cohen, D. A. D., Hyde, M. L., & Prud'homme, M. (1989). A comparison of movement direction–related versus load direction–related activity in primate motor cortex, using a two-dimensional reaching task. *Journal of Neuroscience*, *9*(6), 2080–2102.

Kettner, R. E., Schwartz, A. B., & Georgopoulos, A. P. (1988). Primate motor cortex and free arm movements to visual targets in three-dimensional space. III. Positional gradients and population coding of movement direction from various movement origins. *Journal of Neuroscience*, *8*(8), 2938–2947.

Kiehn, O., & Eken, T. (1997). Prolonged firing in motor units: Evidence of plateau potentials in human motoneurons? *Journal of Neurophysiology*, *78*, 3061–3068.

Lacquaniti, F., Terzuolo, C., & Viviani, P. (1983). The law relating the kinematic and figural aspects of drawing movements. *Acta Psychol.*, *54*, 115–130.

Lee, D., Port, N. L., Kruse, W., & Georgopoulos, A. P. (1998). Variability and correlated noise in the discharge of neurons in motor and parietal areas of the primate cortex. *Journal of Neuroscience*, *18*(3), 1161–1170.

Maltenfort, M. G., Heckman, C. J., & Rymer, Z. W. (1998). Decorrelating actions of Renshaw interneurons on the firing of spinal motoneurons within a motor nucleus: A simulation study. *Journal of Neurophysiology*, *80*, 309–323.

McAuley, J. H., Rothwell, J. C., & Marsden, C. D. (1997). Frequency peaks of tremor, muscle vibration and electromyographic activity at 10 Hz, 20 Hz and 40 Hz during human finger muscle contraction may reflect rythmicities of central neural firing. *Exp. Brain Res.*, *114*, 525–541.

Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E., & Smith, J. E. K. (1988). Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological Review*, *95*, 340–370.

Morasso, P. (1981). Spatial control of arm movements. *Exp. Brain Res.*, *42*, 223–227.

Nakano, E., Imamizu, H., Osu, R., Uno, Y., Gomi, H., Yoshioka, T., & Kawato, M. (1999). Quantitative examinations of internal representations for arm trajectory planning: Minimum commanded torque change model. *Journal of Neurophysiology*, *81*(5), 2140–2155.

Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Biological Cybernetics*, *46*, 135–147.

Pouget, A., Deneve, S., Ducom, J.-C., & Latham, P. (1999). Narrow versus wide tuning curves: What's best for a population code? *Neural Computation*, *11*, 85–90.

Schmidt, R. A., Zelaznik, H., Hawkins, B., Frank, J. S., & Quinn, J. T. J. (1979). Motor-output variability: A theory for the accuracy of rapid motor acts. *Psychological Review*, *86*(5), 415–451.

Snippe, H. (1996). Parameter extraction from population codes: A critical assessment. *Neural Computation*, *8*, 511–529.

Stephens, J. A., Harrison, L. M., Mayston, M. J., Carr, L. J., & Gibbs, J. (1999). The sharing principle. In M. D. Binder (Ed.), *Peripheral and spinal mechanisms in the neural control of movement* (pp. 419–426). Oxford: Elsevier.

Sutton, G. G., & Sykes, K. (1967). The variation of hand tremor with force in healthy subjects. *Journal of Physiology*, *191*(3), 699–711.

Todorov, E. (2000). Direct cortical control of muscle activation in voluntary arm movements: A model. *Nature Neuroscience*, *3*(4), 391–398.

Todorov, E. (2001). Arm movements minimize a combination of error and effort. *Neural Control of Movement, 11*.

Todorov, E., & Jordan, M. I. (1998). Smoothness maximization along a predefined path accurately predicts the speed profiles of complex arm movements. *Journal of Neurophysiology*, *80*, 696–714.

Todorov, E., Li, R., Gandolfo, F., Benda, B., DiLorenzo, D., Padoa-Schioppa, C., & Bizzi, E. (2000). Cosine tuning minimizes motor errors: Theoretical results and experimental confirmation. *Society for Neuroscience Abstracts*, 785.6.

Turner, R. S., Owens, J., & Anderson, M. E. (1995). Directional variation of spatial and temporal characteristics of limb movements made by monkeys in a two-dimensional work space. *Journal of Neurophysiology*, *74*, 684–697.

Winter, D. A. (1990). *Biomechanics and motor control of human movement*. New York: Wiley.

Zajac, F. E. (1989). Muscle and tendon: Properties, models, scaling, and application to biomechanics and motor control. *Critical Reviews in Biomedical Engineering*, *17*(4), 359–411.

Zhang, K., & Sejnowski, T. (1999a). A theory of geometric constraints on neural activity for natural three-dimensional movement. *Journal of Neuroscience*, *19*(8), 3122–3145.

Zhang, K., & Sejnowski, T. J. (1999b). Neuronal tuning: To sharpen or broaden? *Neural Computation*, *11*, 75–84.