Stochastic Optimal Control and Estimation Methods Adapted to the Noise Characteristics of the Sensorimotor System

Emanuel Todorov
todorov@cogsci.ucsd.edu
Department of Cognitive Science, University of California San Diego,
La Jolla CA 92093-0515.

Optimality principles of biological movement are conceptually appealing and straightforward to formulate. Testing them empirically, however, requires the solution to stochastic optimal control and estimation problems for reasonably realistic models of the motor task and the sensorimotor periphery. Recent studies have highlighted the importance of incorporating biologically plausible noise into such models. Here we extend the linear-quadratic-gaussian framework—currently the only framework where such problems can be solved efficiently—to include control-dependent, state-dependent, and internal noise. Under this extended noise model, we derive a coordinate-descent algorithm guaranteed to converge to a feedback control law and a nonadaptive linear estimator optimal with respect to each other. Numerical simulations indicate that convergence is exponential, local minima do not exist, and the restriction to nonadaptive linear estimators has negligible effects in the control problems of interest. The application of the algorithm is illustrated in the context of reaching movements. A Matlab implementation is available at www.cogsci.ucsd.edu/~todorov.

1 Introduction

Many theories in the physical sciences are expressed in terms of optimality principles, which often provide the most compact description of the laws governing a system's behavior. Such principles play an important role in the field of sensorimotor control as well (Todorov, 2004). A quantitative theory of sensorimotor control requires a precise definition of success in the form of a scalar cost function. By combining top-down reasoning with intuitions derived from empirical observations, researchers have proposed a number of hypothetical cost functions for biological movement. While such hypotheses are not difficult to formulate, comparing their predictions to experimental data is complicated by the fact that the predictions have to be derived in the first place—that is, the hypothetical optimal control and estimation problems have to be solved. The most popular approach has been to optimize, in an open loop, the sequence of control signals (Chow & Jacobson,

1971; Hatze & Buys, 1977; Anderson & Pandy, 2001) or limb states (Nelson, 1983; Flash & Hogan, 1985; Uno, Kawato, & Suzuki, 1989; Harris & Wolpert, 1998). For stochastic partially observable plants such as the musculoskeletal system, however, open-loop approaches yield suboptimal performance (Todorov & Jordan, 2002b; Todorov, 2004). Optimal performance can be achieved only by a feedback control law, which uses all sensory data available online to compute the most appropriate muscle activations under the circumstances.

Optimization in the space of feedback control laws is studied in the related fields of stochastic optimal control, dynamic programming, and reinforcement learning. Despite many advances, the general-purpose methods that are guaranteed to converge in a reasonable amount of time to a reasonable answer remain limited to discrete state and action spaces (Bertsekas & Tsitsiklis, 1997; Sutton & Barto, 1998; Kushner & Dupuis, 2001). Discretization methods are well suited for higher-level control problems, such as the problem faced by a rat that has to choose which way to turn in a two-dimensional maze. But the main focus in sensorimotor control is on a different level of analysis: on how the rat chooses a hundred or so graded muscle activations at each point in time, in a way that causes its body to move toward the reward without falling or hitting walls. Even when the musculoskeletal system is idealized and simplified, the state and action spaces of interest remain continuous and high-dimensional, and the curse of dimensionality prevents the use of discretization methods. Generalizations of these methods to continuous high-dimensional spaces typically involve function approximations whose properties are not yet well understood. Such approximations can produce good enough solutions, which is often acceptable in engineering applications. However, the success of a theory of sensorimotor control ultimately depends on its ability to explain data in a principled manner. Unless the theory's predictions are close to the globally optimal solution of the hypothetical control problem, it is difficult to determine whether the (mis)match to experimental data is due to the general (in)applicability of optimality ideas to biological movement, or the (in)appropriateness of the specific cost function, or the specific approximations—in both the plant model and the controller design—used to derive the predictions.

Accelerated progress will require efficient and well-understood methods for optimal feedback control of stochastic, partially observable, continuous, nonstationary, and high-dimensional systems. The only framework that currently provides such methods is linear-quadratic-gaussian (LQG) control, which has been used to model biological systems subject to sensory and motor uncertainty (Loeb, Levine, & He, 1990; Hoff, 1992; Kuo, 1995). While optimal solutions can be obtained efficiently within the LQG setting (via Riccati equations), this computational efficiency comes at the price of reduced biological realism, because (1) musculoskeletal dynamics are generally nonlinear, (2) behaviorally relevant performance criteria are

unlikely to be globally quadratic (Kording & Wolpert, 2004), and (3) noise in the sensorimotor apparatus is not additive but signal-dependent. The third limitation is particularly problematic because it is becoming increasingly clear that many robust and extensively studied phenomena—such as trajectory smoothness, speed-accuracy trade-offs, task-dependent impedance, structured motor variability and synergistic control, and cosine tuning— are linked to the signal-dependent nature of sensorimotor noise (Harris & Wolpert, 1998; Todorov, 2002; Todorov & Jordan, 2002b).

It is thus desirable to extend the LQG setting as much as possible and adapt it to the online control and estimation problems that the nervous system faces. Indeed, extensions are possible in each of the three directions listed above:

1. Nonlinear dynamics (and nonquadratic costs) can be approximated in the vicinity of the expected trajectory generated by an existing controller. One can then apply modified LQG methodology to the approximate problem and use it to improve the existing controller iteratively. Differential dynamic programming (Jacobson & Mayne, 1970), as well as iterative LQG methods (Li & Todorov, 2004; Todorov & Li, 2004), are based on this general idea. In their present form, most such methods assume deterministic dynamics, but stochastic extensions are possible (Todorov & Li, 2004).

2. Quadratic costs can be replaced with a parametric family of exponential-of-quadratic costs, for which optimal LQG-like solutions can be obtained efficiently (Whittle, 1990; Bensoussan, 1992). The controllers that are optimal for such costs range from risk averse (i.e., robust), through classic LQG, to risk seeking. This extended family of cost functions has not yet been explored in the context of biological movement.

3. Additive gaussian noise in the plant dynamics can be replaced with multiplicative noise, which is still gaussian but has standard deviation proportional to the magnitude of the control signals or state variables. When the state of the plant is fully observable, optimal LQG-like solutions can be computed efficiently, as shown by several authors (Kleinman, 1969; McLane, 1971; Willems & Willems, 1976; Bensoussan, 1992; El Ghaoui, 1995; Beghi & D'Alessandro, 1998; Rami, Chen, & Moore, 2001). Such methodology has also been used to model reaching movements (Hoff, 1992). Most relevant to the study of sensorimotor control, however, is the partially observable case, which remains an open problem. While some work along these lines has been done (Pakshin, 1978; Phillis, 1985), it has not produced reliable algorithms that one can use off the shelf in building biologically relevant models (see section 9). Our goal here is to address that problem, and provide the model-building methodology that is needed.

Table 1: List of Notation.

$\mathbf{x}_t \in \mathbb{R}^m$	state vector at time step t
$\mathbf{u}_t \in \mathbb{R}^p$	control signal
$\mathbf{y}_t \in \mathbb{R}^k$	sensory observation
n	total number of time steps
A, B, H	system dynamics and observation matrices
$\boldsymbol{\xi}_t, \boldsymbol{\omega}_t, \boldsymbol{\varepsilon}_t, \boldsymbol{\epsilon}_t, \boldsymbol{\eta}_t$	zero-mean noise terms
$\Omega^{\xi}, \Omega^{\omega}, \Omega^{\varepsilon}, \Omega^{\epsilon}, \Omega^{\eta}$	covariances of noise terms
$C_1, \ldots, C_c$	scaling matrices for control-dependent system noise
$D_1, \ldots, D_d$	scaling matrices for state-dependent observation noise
Q_t, R	matrices defining state- and control-dependent costs
$\widehat{\mathbf{x}}_t$	state estimate
$\mathbf{e}_t$	estimation error
Σ_t	conditional estimation error covariance
$\Sigma_t^{\mathbf{e}}, \Sigma_t^{\widehat{\mathbf{x}}}, \Sigma_t^{\widehat{\mathbf{x}}\mathbf{e}}$	unconditional covariances
v_t	optimal cost-to-go function
$S_t^{\mathbf{x}}, S_t^{\mathbf{e}}, s_t$	parameters of the optimal cost-to-go function
K_t	filter gain matrices
L_t	control gain matrices

In this letter, we define an extended noise model that reflects the properties of the sensorimotor system; derive an efficient algorithm for solving the stochastic optimal control and estimation problems under that noise model; illustrate the application of this extended LQG methodology in the context of reaching movements; and study the properties of the new algorithm through extensive numerical simulations. A special case of the algorithm derived here has already allowed us (Todorov & Jordan, 2002b) to construct models of a wider range of empirical results than previously possible.

In section 2 we motivate our extended noise model, which includes control-dependent, state-dependent, and internal estimation noise. In section 3 we formalize the problem and restrict the feedback control laws under consideration to functions of state estimates that are obtained by unbiased nonadaptive linear filters. In section 4 we compute the optimal feedback control law for any nonadaptive linear filter and show that it is linear in the state estimate. In section 5 we derive the optimal nonadaptive linear filter for any linear control law. The two results together provide an iterative coordinate-descent algorithm (equations 4.2 and 5.2), which is guaranteed to converge to a filter and a control law optimal with respect to each other. In section 6 we illustrate the application of our method to the analysis of reaching movements. In section 7 we explore numerically the convergence properties of the algorithm and observe exponential convergence with no local minima. In section 8 we assess the effects of assuming a nonadaptive linear filter and find them to be negligible for the control problems of interest.

Table 1 shows the notation used in this letter.

2 Noise Characteristics of the Sensorimotor System

Noise in the motor output is not additive but instead increases with the magnitude of the control signals. This is intuitively obvious: if you rest your arm on the table, it does not bounce around (i.e., the passive plant dynamics have little noise), but when you make a movement (i.e., generate control signals), the outcome is not always as desired. Quantitatively, the relationship between motor noise and control magnitude is surprisingly simple. Such noise has been found to be multiplicative: the standard deviation of muscle force is well fit with a linear function of the mean force, in both static (Sutton & Sykes, 1967; Todorov, 2002) and dynamic (Schmidt, Zelaznick, Hawkins, Frank, & Quinn, 1979) isometric force tasks. The exact reasons for this dependence are not entirely clear, although it can be explained at least in part with Poisson noise on the neural level combined with Henneman's size principle of motoneuron recruitment (Jones, Hamilton, & Wolpert, 2002). To formalize the empirically established dependence, let $\mathbf{u}$ be a vector of control signals (corresponding to the muscle activation levels that the nervous system attempts to set) and ε be a vector of zero-mean random numbers. A general multiplicative noise model takes the form $C(\mathbf{u})\varepsilon$, where $C(\mathbf{u})$ is a matrix whose elements depend linearly on $\mathbf{u}$. To express a linear relationship between a vector $\mathbf{u}$ and a matrix C, we make the ith column of C equal to $C_i\mathbf{u}$, where C_i are constant scaling matrices. Then we have $C(\mathbf{u})\varepsilon = \sum_i C_i\mathbf{u}\varepsilon^i$, where ε^i is the ith component of the random vector ε.

Online movement control relies on feedback from a variety of sensory modalities, with vision and proprioception typically playing the dominant role. Visual noise obviously depends on the retinal position of the objects of interest and increases with distance away from the fovea (i.e., eccentricity). The accuracy of visual positional estimates is again surprisingly well modeled with multiplicative noise, whose standard deviation is proportional to eccentricity. This is an instantiation of Weber's law and has been found to be quite robust in a variety of interval discrimination experiments (Burbeck & Yap, 1990; Whitaker & Latham, 1997). We have also confirmed this scaling law in a visuomotor setting, where subjects pointed to memorized targets presented in the visual periphery (Todorov, 1998). Such results motivate the use of a multiplicative observation noise model of the form $D(\mathbf{x})\epsilon = \sum_i D_i\mathbf{x}\epsilon^i$, where $\mathbf{x}$ is the state of the plant and environment, including the current fixation point and the positions and velocities of relevant objects. Incorporating state-dependent noise in analyses of sensorimotor control can allow more accurate modeling of the effects of feedback and various experimental perturbations; it also can effectively induce a cost function over eye movement patterns and allow us to predict the eye movements that would result in optimal hand performance (Todorov, 1998). Note that if other forms of state-dependent sensory noise are found, the model can still be useful as a linear approximation.

Intelligent control of a partially observable stochastic plant requires a feedback control law, which is typically a function of a state estimate that is computed recursively over time. In engineering applications, the estimation-control loop is implemented in a noiseless digital computer, and so all noise is external. In models of biological movement, we usually make the same assumption, treating all noise as being a property of the musculoskeletal plant or the sensory apparatus. This is in principle unrealistic, because neural representations are likely subject to internal fluctuations that do not arise in the periphery. It is also unrealistic in modeling practice. An ideal observer model predicts that the estimation error covariance of any stationary feature of the environment will asymptote to 0. In particular, such models predict that if we view a stationary object in the visual periphery long enough, we should eventually know exactly where it is and be able to reach for it as accurately as if it were at the center of fixation. This contradicts our intuition as well as experimental data. Both interval discrimination experiments and reaching to remembered peripheral targets experiments indicate that estimation errors asymptote rather quickly, but not to 0. Instead, the asymptote level depends linearly on eccentricity. The simplest way to model this is to assume another noise process, which we call internal noise, acting directly on whatever state estimate the nervous system chooses to compute.

3 Problem Statement and Assumptions

Consider a linear dynamical system with state $\mathbf{x}_t \in \mathbb{R}^m$, control $\mathbf{u}_t \in \mathbb{R}^p$, feedback $\mathbf{y}_t \in \mathbb{R}^k$, in discrete time t:

$$
\begin{aligned}
\textit{Dynamics} \qquad & \mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t + \boldsymbol{\xi}_t + \sum_{i=1}^{c} \varepsilon_t^i C_i \mathbf{u}_t \\
\textit{Feedback} \qquad & \mathbf{y}_t = H\mathbf{x}_t + \boldsymbol{\omega}_t + \sum_{i=1}^{d} \epsilon_t^i D_i \mathbf{x}_t \\
\textit{Cost per step} \quad & \mathbf{x}_t^T Q_t \mathbf{x}_t + \mathbf{u}_t^T R \mathbf{u}_t
\end{aligned}
\qquad (3.1)
$$

The feedback signal $\mathbf{y}_t$ is received after the control signal $\mathbf{u}_t$ has been generated. The initial state has known mean $\widehat{\mathbf{x}}_1$ and covariance Σ_1. All matrices are known and have compatible dimensions; making them time varying is straightforward. The control cost matrix R is symmetric positive definite ($R > 0$), and the state cost matrices $Q_1, \ldots, Q_n$ are symmetric positive semidefinite ($Q_t \geq 0$). Each movement lasts n time steps; at $t = n$, the final cost is $\mathbf{x}_n^T Q_n \mathbf{x}_n$, and $\mathbf{u}_n$ is undefined. The independent random variables $\boldsymbol{\xi}_t \in \mathbb{R}^m$, $\boldsymbol{\omega}_t \in \mathbb{R}^k$, $\varepsilon_t \in \mathbb{R}^c$, and $\epsilon_t \in \mathbb{R}^d$ have multidimensional gaussian distributions with mean 0 and covariances $\Omega^\xi \geq 0$, $\Omega^\omega > 0$, $\Omega^\varepsilon = I$ and $\Omega^\epsilon = I$ respectively. Thus, the control-dependent and state-dependent noise terms in equation 3.1 have covariances $\sum_i C_i \mathbf{u}_t \mathbf{u}_t^T C_i^T$ and $\sum_i D_i \mathbf{x}_t \mathbf{x}_t^T D_i^T$. When the

control-dependent noise is meant to be added to the control signal (which is usually the case), the matrices C_i should have the form BF_i where F_i are the actual noise scaling factors. Then the control-dependent part of the plant dynamics becomes $B(I + \sum_i \varepsilon_t^i F_i)\mathbf{u}_t$.

The problem of optimal control is to find the optimal control law, that is, the sequence of causal control functions $\mathbf{u}_t(\mathbf{u}_1, \ldots, \mathbf{u}_{t-1}, \mathbf{y}_1, \ldots, \mathbf{y}_{t-1})$ that minimize the expected total cost over the movement. Note that computing the optimal sequence of functions $\mathbf{u}_1(\cdot), \ldots, \mathbf{u}_{n-1}(\cdot)$ is a different, and in general much more difficult, problem than computing the optimal sequence of open-loop controls $\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$.

When only additive noise is present (i.e., $C_1, \ldots, C_c = 0$ and $D_1, \ldots, D_d = 0$), this reduces to the classic LQG problem, which has the well-known optimal solution (Davis & Vinter, 1985)

Linear-Quadratic Regulator   *Kalman Filter*

$$\mathbf{u}_t = -L_t \widehat{\mathbf{x}}_t \qquad\qquad \widehat{\mathbf{x}}_{t+1} = A\widehat{\mathbf{x}}_t + B\mathbf{u}_t + K_t(\mathbf{y}_t - H\widehat{\mathbf{x}}_t)$$

$$L_t = (R + B^\mathsf{T} S_{t+1} B)^{-1} B^\mathsf{T} S_{t+1} A \qquad K_t = A\Sigma_t H^\mathsf{T} (H\Sigma_t H^\mathsf{T} + \Omega^\omega)^{-1} \qquad (3.2)$$

$$S_t = Q_t + A^\mathsf{T} S_{t+1}(A - BL_t) \qquad \Sigma_{t+1} = \Omega^\xi + (A - K_t H)\Sigma_t A^\mathsf{T}$$

In that case, the optimal control law depends on the history of control and feedback signals only through the state estimate $\widehat{\mathbf{x}}_t$, which is updated recursively by the Kalman filter. The matrices L that define the optimal control law do not depend on the noise covariances or filter coefficients, and the matrices K that define the optimal filter do not depend on the cost and control law.

In the case of control-dependent and state-dependent noise, the above independence properties no longer hold. This complicates the problem substantially and forces us to adopt a more restricted formulation in the interest of analytical tractability. We assume that, as in equation 3.2, the entire history of control and feedback signals is summarized by a state estimate $\widehat{\mathbf{x}}_t$, which is all the information available to the control system at time t. The feedback control law $\mathbf{u}_t(\cdot)$ is allowed to be an arbitrary function of $\widehat{\mathbf{x}}_t$, but $\widehat{\mathbf{x}}_t$ can be updated only by a recursive linear filter of the form:

$$\widehat{\mathbf{x}}_{t+1} = A\widehat{\mathbf{x}}_t + B\mathbf{u}_t + K_t(\mathbf{y}_t - H\widehat{\mathbf{x}}_t) + \eta_t.$$

The internal noise $\eta_t \in \mathbb{R}^m$ has mean 0 and covariance $\Omega^\eta \geq 0$. The filter gains $K_1, \ldots, K_{n-1}$ are nonadaptive; they are determined in advance and cannot change as a function of the specific controls and observations within a simulation run. Such a filter is always unbiased: for any $K_1, \ldots, K_{n-1}$, we have $\mathrm{E}[\mathbf{x}_t|\widehat{\mathbf{x}}_t] = \widehat{\mathbf{x}}_t$ for all t. Note, however, that under the extended noise model, any nonadaptive linear filter is suboptimal: when $\widehat{\mathbf{x}}_t$ is computed as defined above, $\mathrm{Cov}[\mathbf{x}_t|\widehat{\mathbf{x}}_t]$ is generally larger than $\mathrm{Cov}[\mathbf{x}_t|\mathbf{u}_1, \ldots, \mathbf{u}_{t-1}, \mathbf{y}_1, \ldots, \mathbf{y}_{t-1}]$. The consequences of this will be explored numerically in section 8.

4 Optimal Controller

The optimal $\mathbf{u}_t$ will be computed using the method of dynamic programming. We will show by induction that if the true state at time t is $\mathbf{x}_t$ and the unbiased state estimate available to the control system is $\widehat{\mathbf{x}}_t$, then the optimal cost-to-go function (i.e., the cost expected to accumulate under the optimal control law) has the quadratic form

$$v_t(\mathbf{x}_t, \widehat{\mathbf{x}}_t) = \mathbf{x}_t^\mathsf{T} S_t^\mathsf{x} \mathbf{x}_t + (\mathbf{x}_t - \widehat{\mathbf{x}}_t)^\mathsf{T} S_t^\mathsf{e}(\mathbf{x}_t - \widehat{\mathbf{x}}_t) + s_t = \mathbf{x}_t^\mathsf{T} S_t^\mathsf{x} \mathbf{x}_t + \mathbf{e}_t^\mathsf{T} S_t^\mathsf{e} \mathbf{e}_t + s_t,$$

where $\mathbf{e}_t \triangleq \mathbf{x}_t - \widehat{\mathbf{x}}_t$ is the estimation error. At the final time $t = n$, the optimal cost-to-go is simply the final cost $\mathbf{x}_n^\mathsf{T} Q_n \mathbf{x}_n$, and so v_n is in the assumed form with $S_n^\mathsf{x} = Q_n$, $S_n^\mathsf{e} = 0$, $s_n = 0$. To carry out the induction proof, we have to show that if v_{t+1} is in the above form for some $t < n$, then v_t is also in that form.

Consider a time-varying control law that is optimal at times $t + 1, \ldots, n$, and at time t is given by $\mathbf{u}_t = \pi(\widehat{\mathbf{x}}_t)$. Let $v_t^\pi(\mathbf{x}_t, \widehat{\mathbf{x}}_t)$ be the corresponding cost-to-go function. Since this control law is optimal after time t, we have $v_{t+1}^\pi = v_{t+1}$. Then the cost-to-go function v_t^π satisfies the Bellman equation:

$$v_t^\pi(\mathbf{x}_t, \widehat{\mathbf{x}}_t) = \mathbf{x}_t^\mathsf{T} Q_t \mathbf{x}_t + \pi(\widehat{\mathbf{x}}_t)^\mathsf{T} R \pi(\widehat{\mathbf{x}}_t) + \mathrm{E}\left[v_{t+1}(\mathbf{x}_{t+1}, \widehat{\mathbf{x}}_{t+1}) | \mathbf{x}_t, \widehat{\mathbf{x}}_t, \pi\right].$$

To compute the above expectation term, we need the update equations for the system variables. Using the definitions of the observation $\mathbf{y}_t$ and the estimation error $\mathbf{e}_t$, the stochastic dynamics of the variables of interest become

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + B\pi(\widehat{\mathbf{x}}_t) + \boldsymbol{\xi}_t + \sum_i \varepsilon_t^i C_i \pi(\widehat{\mathbf{x}}_t)$$

$$\mathbf{e}_{t+1} = (A - K_t H)\mathbf{e}_t + \boldsymbol{\xi}_t - K_t \boldsymbol{\omega}_t - \boldsymbol{\eta}_t + \sum_i \varepsilon_t^i C_i \pi(\widehat{\mathbf{x}}_t) - \sum_i \epsilon_t^i K_t D_i \mathbf{x}_t.$$
$$(4.1)$$

Then the conditional means and covariances of $\mathbf{x}_{t+1}$ and $\mathbf{e}_{t+1}$ are

$$\mathrm{E}\left[\mathbf{x}_{t+1} | \mathbf{x}_t, \widehat{\mathbf{x}}_t, \pi\right] = A\mathbf{x}_t + B\pi(\widehat{\mathbf{x}}_t)$$

$$\mathrm{E}\left[\mathbf{e}_{t+1} | \mathbf{x}_t, \widehat{\mathbf{x}}_t, \pi\right] = (A - K_t H)\mathbf{e}_t$$

$$\mathrm{Cov}\left[\mathbf{x}_{t+1} | \mathbf{x}_t, \widehat{\mathbf{x}}_t, \pi\right] = \Omega^\xi + \sum_i C_i \pi(\widehat{\mathbf{x}}_t)\pi(\widehat{\mathbf{x}}_t)^\mathsf{T} C_i^\mathsf{T}$$

$$\mathrm{Cov}\left[\mathbf{e}_{t+1} | \mathbf{x}_t, \widehat{\mathbf{x}}_t, \pi\right] = \Omega^\xi + \sum_i C_i \pi(\widehat{\mathbf{x}}_t)\pi(\widehat{\mathbf{x}}_t)^\mathsf{T} C_i^\mathsf{T} + \Omega^\eta$$

$$+ K_t \Omega^\omega K_t^\mathsf{T} + \sum_i K_t D_i \mathbf{x}_t \mathbf{x}_t^\mathsf{T} D_i^\mathsf{T} K_t^\mathsf{T},$$

and the conditional expectation in the Bellman equation can be computed. The cost-to-go becomes

$$v_t^\pi(\mathbf{x}_t, \widehat{\mathbf{x}}_t) = \mathbf{x}_t^\mathsf{T}\big(Q_t + A^\mathsf{T} S_{t+1}^\mathsf{x} A + \mathcal{D}_t\big)\mathbf{x}_t$$
$$+ \mathbf{e}_t^\mathsf{T}(A - K_t H)^\mathsf{T} S_{t+1}^\mathsf{e}(A - K_t H)\mathbf{e}_t$$
$$+ \operatorname{tr}(\mathcal{M}_t) + \pi\,(\widehat{\mathbf{x}}_t)^\mathsf{T}\big(R + B^\mathsf{T} S_{t+1}^\mathsf{x} B + \mathcal{C}_t\big)\pi(\widehat{\mathbf{x}}_t)$$
$$+ 2\pi(\widehat{\mathbf{x}}_t)^\mathsf{T} B^\mathsf{T} S_{t+1}^\mathsf{x} A\mathbf{x}_t,$$

where we defined the shortcuts

$$\mathcal{C}_t \triangleq \sum_i C_i^\mathsf{T}\big(S_{t+1}^\mathsf{e} + S_{t+1}^\mathsf{x}\big) C_i,$$

$$\mathcal{D}_t \triangleq \sum_i D_i^\mathsf{T} K_t^\mathsf{T} S_{t+1}^\mathsf{e} K_t D_i, \quad \text{and}$$

$$\mathcal{M}_t \triangleq S_{t+1}^\mathsf{x}\Omega^\xi + S_{t+1}^\mathsf{e}\big(\Omega^\xi + \Omega^\eta + K_t \Omega^\omega K_t^\mathsf{T}\big).$$

Note that the control law affects only the cost-go-to function through an expression that is quadratic in $\pi(\widehat{\mathbf{x}}_t)$, which can be minimized analytically. But there is a problem: the minimum depends on $\mathbf{x}_t$, while π is only allowed to be a function of $\widehat{\mathbf{x}}_t$. To obtain the optimal control law at time t, we have to take an expectation over $\mathbf{x}_t$ conditional on $\widehat{\mathbf{x}}_t$, and find the function π that minimizes the resulting expression. Note that the control-dependent expression is linear in $\mathbf{x}_t$, and so its expectation depends on the conditional mean of $\mathbf{x}_t$ but not on any higher moments. Since $\mathrm{E}\,[\mathbf{x}_t|\widehat{\mathbf{x}}_t] = \widehat{\mathbf{x}}_t$, we have

$$\mathrm{E}\big[v_t^\pi(\mathbf{x}_t, \widehat{\mathbf{x}}_t)|\widehat{\mathbf{x}}_t\big] = \text{const} + \pi(\widehat{\mathbf{x}}_t)^\mathsf{T}\big(R + B^\mathsf{T} S_{t+1}^\mathsf{x} B + \mathcal{C}_t\big)\pi(\widehat{\mathbf{x}}_t)$$
$$+ 2\pi(\widehat{\mathbf{x}}_t)^\mathsf{T} B^\mathsf{T} S_{t+1}^\mathsf{x} A\widehat{\mathbf{x}}_t,$$

and thus the optimal control law at time t is

$$\mathbf{u}_t = \pi(\widehat{\mathbf{x}}_t) = -L_t\widehat{\mathbf{x}}_t; \qquad L_t \triangleq \big(R + B^\mathsf{T} S_{t+1}^\mathsf{x} B + \mathcal{C}_t\big)^{-1} B^\mathsf{T} S_{t+1}^\mathsf{x} A.$$

Note that the linear form of the optimal control law fell out of the optimization and was not assumed. Given our assumptions, the matrix being inverted is symmetric positive-definite.

To complete the induction proof, we have to compute the optimal cost-to-go v_t, which is equal to v_t^π when π is set to the optimal control law $-L_t\widehat{\mathbf{x}}_t$. Using the fact that $L_t^\mathsf{T}(R + B^\mathsf{T} S_{t+1}^\mathsf{x} B + \mathcal{C}_t)L_t = L_t^\mathsf{T} B^\mathsf{T} S_{t+1}^\mathsf{x} A = A^\mathsf{T} S_{t+1}^\mathsf{x} B L_t$, and that $\widehat{\mathbf{x}}^\mathsf{T} Z\widehat{\mathbf{x}} - 2\widehat{\mathbf{x}}^\mathsf{T} Z\mathbf{x} = (\mathbf{x} - \widehat{\mathbf{x}})^\mathsf{T} Z(\mathbf{x} - \widehat{\mathbf{x}}) - \mathbf{x}^\mathsf{T} Z\mathbf{x} = \mathbf{e}^\mathsf{T} Z\mathbf{e} - \mathbf{x}^\mathsf{T} Z\mathbf{x}$ for a symmetric matrix Z (in our case equal to $L_t^\mathsf{T} B^\mathsf{T} S_{t+1}^\mathsf{x} A$), the result is

$$v_t(\mathbf{x}_t, \widehat{\mathbf{x}}_t) = \mathbf{x}_t^\mathsf{T}\big(Q_t + A^\mathsf{T} S_{t+1}^\mathsf{x}(A - BL_t) + \mathcal{D}_t\big)\mathbf{x}_t + \operatorname{tr}(\mathcal{M}_t) + s_{t+1}$$
$$+ \mathbf{e}_t^\mathsf{T}\big(A^\mathsf{T} S_{t+1}^\mathsf{x} BL_t + (A - K_t H)^\mathsf{T} S_{t+1}^\mathsf{e}(A - K_t H)\big)\mathbf{e}_t.$$

We now see that the optimal cost-to-go function remains in the assumed quadratic form, which completes the induction proof. The optimal control law is computed recursively backward in time as

$\textit{Controller} \quad \mathbf{u}_t = -L_t \widehat{\mathbf{x}}_t$

$$L_t = \left(R + B^\mathsf{T} S_{t+1}^{\mathsf{x}} B + \sum_i C_i^\mathsf{T} (S_{t+1}^{\mathsf{x}} + S_{t+1}^{\mathsf{e}}) C_i \right)^{-1} B^\mathsf{T} S_{t+1}^{\mathsf{x}} A$$

$$S_t^{\mathsf{x}} = Q_t + A^\mathsf{T} S_{t+1}^{\mathsf{x}} (A - BL_t) + \sum_i D_i^\mathsf{T} K_t^\mathsf{T} S_{t+1}^{\mathsf{e}} K_t D_i; \qquad S_n^{\mathsf{x}} = Q_n \qquad (4.2)$$

$$S_t^{\mathsf{e}} = A^\mathsf{T} S_{t+1}^{\mathsf{x}} BL_t + (A - K_t H)^\mathsf{T} S_{t+1}^{\mathsf{e}} (A - K_t H); \qquad S_n^{\mathsf{e}} = 0$$

$$s_t = \mathrm{tr}\big(S_{t+1}^{\mathsf{x}} \Omega^\xi + S_{t+1}^{\mathsf{e}} \big(\Omega^\xi + \Omega^\eta + K_t \Omega^\omega K_t^\mathsf{T}\big)\big) + s_{t+1}; \qquad s_n = 0.$$

The total expected cost is $\widehat{\mathbf{x}}_1^\mathsf{T} S_1^{\mathsf{x}} \widehat{\mathbf{x}}_1 + \mathrm{tr}((S_1^{\mathsf{x}} + S_1^{\mathsf{e}})\Sigma_1) + s_1$.

When the control-dependent and state-dependent noise terms are removed (i.e., $C_1, \ldots, C_c = 0$, $D_1, \ldots, D_d = 0$), the control laws given by equation 4.2 and 3.2 are identical. The internal noise term η, as well as the additive noise terms ξ and ω, do not directly affect the calculation of the feedback gain matrices L. However, all noise terms affect the calculation (see below) of the optimal filter gains K, which in turn affect L.

One can attempt to transform equation 3.1 into a fully observable system by setting $H = I$, $\Omega^\omega = \Omega^\eta = 0$, $D_1, \ldots, D_d = 0$, in which case $K = A$, and apply equation 4.2. Recall, however, our assumption that the control signal is generated before the current state is measured. Thus, even if we make the sensory measurement equal to the state, we would still be dealing with a partially observable system. To derive the optimal controller for the fully observable case, we have to assume that $\mathbf{x}_t$ is known at the time when $\mathbf{u}_t$ is generated. The above derivation is now much simplified: the optimal cost-to-go function v_t is in the form $\mathbf{x}_t^\mathsf{T} S_t \mathbf{x}_t + s_t$, and the expectation term that needs to be minimized with regard to $\mathbf{u}_t = \pi(\mathbf{x}_t)$ becomes

$$\mathrm{E}\,[v_{t+1}] = (A\mathbf{x}_t + B\mathbf{u}_t)^\mathsf{T} S_{t+1} (A\mathbf{x}_t + B\mathbf{u}_t)$$

$$+ \mathbf{u}_t^\mathsf{T} \left(\sum_i C_i^\mathsf{T} S_{t+1} C_i \right) \mathbf{u}_t + \mathrm{tr}\,[S_{t+1}\Omega^\xi] + s_{t+1},$$

and the optimal controller is computed in a backward pass through time as

$\textit{Fully observable controller} \quad \mathbf{u}_t = -L_t \mathbf{x}_t$

$$L_t = \left(R + B^\mathsf{T} S_{t+1} B + \sum_i C_i^\mathsf{T} S_{t+1} C_i \right)^{-1} B^\mathsf{T} S_{t+1} A$$

$$S_t = Q_t + A^\mathsf{T} S_{t+1} (A - BL_t); \qquad S_n = Q_n \qquad (4.3)$$

$$s_t = \mathrm{tr}\,(S_{t+1}\Omega^\xi) + s_{t+1}; \qquad s_n = 0.$$

5 Optimal Estimator

So far, we have computed the optimal control law L for any fixed sequence of filter gains K. What should these gains be fixed to? Ideally they should correspond to a Kalman filter, which is the optimal linear estimator. However, in the presence of control-dependent and state-dependent noise, the Kalman filter gains become adaptive (i.e., K_t depends on $\widehat{\mathbf{x}}_t$ and $\mathbf{u}_t$), which would make our control law derivation invalid. Thus, if we want to preserve the optimality of the control law given by equation 4.2 and obtain an iterative algorithm with guaranteed convergence, we need to compute a fixed sequence of filter gains that are optimal for a given control law. Once the iterative algorithm has converged and the control law has been designed, we could use an adaptive filter in place of the fixed-gain filter in run time (see section 8).

Thus, our objective here is the following: given a linear feedback control law $L_1, \ldots, L_{n-1}$ (which is optimal for the previous filter $K_1, \ldots, K_{n-1}$), compute a new filter that, in conjunction with the given control law, results in minimal expected cost. In other words, we will evaluate the filter not by the magnitude of its estimation errors, but by the effect that these estimation errors have on the performance of the composite estimation-control system.

We will show that the new optimal filter can be designed in a forward pass through time. In particular, we will show that regardless of the new values of $K_1, \ldots, K_{t-1}$, the optimal K_t can be found analytically as long as $K_{t+1}, \ldots, K_{n-1}$ still have the values for which $L_{t+1}, \ldots, L_{n-1}$ are optimal. Recall that the optimal $L_{t+1}, \ldots, L_{n-1}$ depend only on $K_{t+1}, \ldots, K_{n-1}$, and so the parameters (as well as the form) of the optimal cost-to-go function v_{t+1} cannot be affected by changing $K_1, \ldots, K_t$. Since K_t affects only the computation of $\widehat{\mathbf{x}}_{t+1}$, and the effect of $\widehat{\mathbf{x}}_{t+1}$ on the total expected cost is captured by the function v_{t+1}, we have to minimize v_{t+1} with respect to K_t. But v is a function of $\mathbf{x}$ and $\widehat{\mathbf{x}}$, while K cannot be adapted to the specific values of $\mathbf{x}$ and $\widehat{\mathbf{x}}$ within a simulation run (by assumption). Thus, the quantity we have to minimize is the unconditional expectation of v_{t+1}. In doing so, we will use that fact that

$$\mathrm{E}\,[v_{t+1}(\mathbf{x}_{t+1}, \widehat{\mathbf{x}}_{t+1})] = \mathrm{E}_{\mathbf{x}_t, \widehat{\mathbf{x}}_t}[\mathrm{E}\,[v_{t+1}(\mathbf{x}_{t+1}, \widehat{\mathbf{x}}_{t+1})|\mathbf{x}_t, \widehat{\mathbf{x}}_t, L_t]].$$

The conditional expectation was already computed as an intermediate step in the previous section (not shown). The terms in $\mathrm{E}\,[v_{t+1}(\mathbf{x}_{t+1}, \widehat{\mathbf{x}}_{t+1})|\mathbf{x}_t, \widehat{\mathbf{x}}_t, L_t]$ that depend on K_t are

$$\mathbf{e}_t^{\mathsf{T}}(A - K_t H)^{\mathsf{T}} S_{t+1}^{\mathbf{e}} (A - K_t H)\mathbf{e}_t + \mathrm{tr}\left(K_t \left(\Omega^\omega + \sum_i D_i \mathbf{x}_t \mathbf{x}_t^{\mathsf{T}} D_i^{\mathsf{T}}\right) K_t^{\mathsf{T}} S_{t+1}^{\mathbf{e}}\right).$$

Defining the (uncentered) unconditional covariances $\Sigma_t^{\mathbf{e}} \triangleq \mathrm{E}\,[\mathbf{e}_t \mathbf{e}_t^{\mathsf{T}}]$ and $\Sigma_t^{\mathbf{x}} \triangleq \mathrm{E}\,[\mathbf{x}_t \mathbf{x}_t^{\mathsf{T}}]$, the unconditional expectation of the K_t-dependent expression

above becomes

$$a(K_t) = \text{tr}\big(\big((A - K_t H)\Sigma_t^e(A - K_t H)^\mathsf{T} + K_t \mathcal{P}_t K_t^\mathsf{T}\big)S_{t+1}^e\big);$$

$$\mathcal{P}_t \triangleq \Omega^\omega + \sum_i D_i \Sigma_t^x D_i^\mathsf{T}.$$

The minimum of $a\,(K_t)$ is found by setting its derivative with regard to K_t to 0. Using the matrix identities $\frac{\partial}{\partial X}\text{tr}\,(XU) = U^\mathsf{T}$ and $\frac{\partial}{\partial X}\text{tr}\,\big(XUX^\mathsf{T}V\big) = VXU + V^\mathsf{T}XU^\mathsf{T}$, and the fact that the matrices S_{t+1}^e, Ω^ω, Σ_t^e, Σ_t^x are symmetric, we obtain

$$\frac{\partial a\,(K_t)}{\partial K_t} = 2S_{t+1}^e\big(K_t\big(H\Sigma_t^e H^\mathsf{T} + \mathcal{P}_t\big) - A\Sigma_t^e H^\mathsf{T}\big).$$

This expression is equal to 0 whenever $K_t = A\Sigma_t^e H^\mathsf{T}(H\Sigma_t^e H^\mathsf{T} + \mathcal{P}_t)^{-1}$, regardless of the value of S_{t+1}^e. Given our assumptions, the matrix being inverted is symmetric positive-definite. Note that the optimal K_t depends on $K_1, \ldots, K_{t-1}$ (through Σ_t^e and Σ_t^x) but is independent of $K_{t+1}, \ldots, K_{n-1}$ (since it is independent of S_{t+1}^e). This is the reason that the filter gains are reoptimized in a forward pass.

To complete the derivation, we have to substitute the optimal filter gains and compute the unconditional covariances. Recall that the variables $\mathbf{x}_t, \widehat{\mathbf{x}}_t, \mathbf{e}_t$ are deterministically related by $\mathbf{e}_t = \mathbf{x}_t - \widehat{\mathbf{x}}_t$, so the covariance of any one of them can be computed given the covariances of the other two, and we have a choice of which pair of covariance matrices to compute. The resulting equations are most compact for the pair $\widehat{\mathbf{x}}_t, \mathbf{e}_t$. The stochastic dynamics of these variables are

$$\widehat{\mathbf{x}}_{t+1} = (A - BL_t)\widehat{\mathbf{x}}_t + K_t H \mathbf{e}_t + K_t \omega_t + \eta_t + \sum_i \epsilon_t^i K_t D_i(\mathbf{e}_t + \widehat{\mathbf{x}}_t).$$

$$\mathbf{e}_{t+1} = (A - K_t H)\mathbf{e}_t + \xi_t - K_t \omega_t - \eta_t - \sum_i \varepsilon_t^i C_i L_t \widehat{\mathbf{x}}_t \qquad (5.1)$$

$$- \sum_i \epsilon_t^i K_t D_i(\mathbf{e}_t + \widehat{\mathbf{x}}_t).$$

Define the unconditional covariances,

$$\Sigma_t^e \triangleq \text{E}\big[\mathbf{e}_t \mathbf{e}_t^\mathsf{T}\big]; \qquad \Sigma_t^{\widehat{x}} \triangleq \text{E}\big[\widehat{\mathbf{x}}_t \widehat{\mathbf{x}}_t^\mathsf{T}\big]; \qquad \Sigma_t^{\widehat{x}e} \triangleq \text{E}\big[\widehat{\mathbf{x}}_t \mathbf{e}_t^\mathsf{T}\big],$$

noting that $\Sigma_t^{\widehat{x}}$ is uncentered and $\Sigma_t^{e\widehat{x}} = (\Sigma_t^{\widehat{x}e})^\mathsf{T}$. Since $\widehat{\mathbf{x}}_1$ is a known constant, the initialization at $t = 1$ is $\Sigma_1^e = \Sigma_1$, $\Sigma_1^{\widehat{x}} = \widehat{\mathbf{x}}_1 \widehat{\mathbf{x}}_1^\mathsf{T}$, $\Sigma_1^{\widehat{x}e} = 0$. With these definitions, we have $\Sigma_t^x = \text{E}\,[(\mathbf{e}_t + \widehat{\mathbf{x}}_t)(\mathbf{e}_t + \widehat{\mathbf{x}}_t)^\mathsf{T}] = \Sigma_t^e + \Sigma_t^{\widehat{x}} + \Sigma_t^{\widehat{x}e} + \Sigma_t^{\widehat{x}e^\mathsf{T}}$. Using equation 5.1, the updates for the unconditional covariances are

$$\Sigma_{t+1}^e = (A - K_t H)\Sigma_t^e(A - K_t H)^\mathsf{T} + \Omega^\xi + \Omega^\eta + K_t \mathcal{P}_t K_t^\mathsf{T}$$

$$+ \sum_i C_i L_t \Sigma_t^{\widehat{x}} L_t^\mathsf{T} C_i^\mathsf{T}$$

$$\Sigma_{t+1}^{\widehat{x}} = (A - BL_t)\Sigma_t^{\widehat{x}}(A - BL_t)^\mathsf{T} + \Omega^\eta + K_t\big(H\Sigma_t^{\mathbf{e}}H^\mathsf{T} + \mathcal{P}_t\big)K_t^\mathsf{T}$$
$$+ (A - BL_t)\Sigma_t^{\widehat{x}\mathbf{e}}H^\mathsf{T}K_t^\mathsf{T} + K_t H\Sigma_t^{\mathbf{e}\widehat{x}}(A - BL_t)^\mathsf{T}$$
$$\Sigma_{t+1}^{\widehat{x}\mathbf{e}} = (A - BL_t)\Sigma_t^{\widehat{x}\mathbf{e}}(A - K_t H)^\mathsf{T} + K_t H\Sigma_t^{\mathbf{e}}(A - K_t H)^\mathsf{T}$$
$$- \Omega^\eta - K_t\mathcal{P}_t K_t^\mathsf{T}.$$

Substituting the optimal value of K_t, which allows some simplifications to the above update equations, the optimal nonadaptive linear filter is computed in a forward pass through time as

$$\textit{Estimator} \quad \widehat{\mathbf{x}}_{t+1} = (A - BL_t)\widehat{\mathbf{x}}_t + K_t(\mathbf{y}_t - H\widehat{\mathbf{x}}_t) + \boldsymbol{\eta}_t$$

$$K_t = A\Sigma_t^{\mathbf{e}}H^\mathsf{T}\left(H\Sigma_t^{\mathbf{e}}H^\mathsf{T} + \Omega^\omega + \sum_i D_i\big(\Sigma_t^{\mathbf{e}} + \Sigma_t^{\widehat{x}} + \Sigma_t^{\widehat{x}\mathbf{e}} + \Sigma_t^{\mathbf{e}\widehat{x}}\big)D_i^\mathsf{T}\right)^{-1}$$

$$\Sigma_{t+1}^{\mathbf{e}} = \Omega^\xi + \Omega^\eta + (A - K_t H)\Sigma_t^{\mathbf{e}}A^\mathsf{T} + \sum_i C_i L_t \Sigma_t^{\widehat{x}}L_t^\mathsf{T}C_i^\mathsf{T}; \qquad \Sigma_1^{\mathbf{e}} = \Sigma_1$$

$$\Sigma_{t+1}^{\widehat{x}} = \Omega^\eta + K_t H\Sigma_t^{\mathbf{e}}A^\mathsf{T} + (A - BL_t)\Sigma_t^{\widehat{x}}(A - BL_t)^\mathsf{T}$$
$$+ (A - BL_t)\Sigma_t^{\widehat{x}\mathbf{e}}H^\mathsf{T}K_t^\mathsf{T} + K_t H\Sigma_t^{\mathbf{e}\widehat{x}}(A - BL_t)^\mathsf{T}; \qquad \Sigma_1^{\widehat{x}} = \widehat{\mathbf{x}}_1\widehat{\mathbf{x}}_1^\mathsf{T}$$
$$\Sigma_{t+1}^{\widehat{x}\mathbf{e}} = (A - BL_t)\Sigma_t^{\widehat{x}\mathbf{e}}(A - K_t H)^\mathsf{T} - \Omega^\eta; \qquad \Sigma_1^{\widehat{x}\mathbf{e}} = 0.$$

$$(5.2)$$

It is worth noting the effects of the internal noise $\boldsymbol{\eta}_t$. If that term did not exist (i.e., $\Omega^\eta = 0$), the last update equation would yield $\Sigma_t^{\widehat{x}\mathbf{e}} = 0$ for all t. Indeed, for an optimal filter, one would expect $\Sigma_t^{\widehat{x}\mathbf{e}} = 0$ from the orthogonality principle: if the state estimate and estimation error were correlated, one could improve the filter by taking that correlation into account. However, the situation here is different because we have noise acting directly on the state estimate. When such noise pushes $\widehat{\mathbf{x}}_t$ in one direction, $\mathbf{e}_t$ is (by definition) pushed in the opposite direction, creating a negative correlation between $\widehat{\mathbf{x}}_t$ and $\mathbf{e}_t$. This is the reason for the negative sign in front of the Ω^η term in the last update equation.

The complete algorithm is the following:

Algorithm:

Initialize $K_1, \ldots, K_{n-1}$, and iterate equation 4.2 and equation 5.2 until convergence. Convergence is guaranteed, because the expected cost is nonnegative by definition, and we are using a coordinate-descent algorithm, which decreases the expected cost in each step. The initial sequence K could be set to 0—in which case, the first pass of equation 4.2 will find the optimal open-loop controls, or initialized from equation 3.2—which is equivalent to assuming additive noise in the first pass.

We can also derive the optimal adaptive linear filter, with gains K_t that depend on the specific $\widehat{\mathbf{x}}_t$ and $\mathbf{u}_t = -L_t\widehat{\mathbf{x}}_t$ within each simulation run. This is again accomplished by minimizing $\mathrm{E}[v_{t+1}]$ with respect to K_t, but the expectation is computed with $\widehat{\mathbf{x}}_t$ being a known constant rather than a random variable. We now have $\Sigma_t^{\widehat{x}} = \widehat{\mathbf{x}}_t\widehat{\mathbf{x}}_t^\mathsf{T}$ and $\Sigma_t^{\widehat{x}\mathbf{e}} = 0$, and so the last two update

equations in equation 5.2 are no longer needed. The optimal adaptive linear filter is

$$\text{Adaptive estimator} \quad \widehat{\mathbf{x}}_{t+1} = (A - BL_t)\widehat{\mathbf{x}}_t + K_t(\mathbf{y}_t - H\widehat{\mathbf{x}}_t) + \eta_t$$

$$K_t = A\Sigma_t H^{\mathsf{T}}\left(H\Sigma_t H^{\mathsf{T}} + \Omega^\omega + \sum_i D_i(\Sigma_t + \widehat{\mathbf{x}}_t\widehat{\mathbf{x}}_t^{\mathsf{T}})D_i^{\mathsf{T}}\right)^{-1} \quad (5.3)$$

$$\Sigma_{t+1} = \Omega^\xi + \Omega^\eta + (A - K_t H)\Sigma_t A^{\mathsf{T}} + \sum_i C_i L_t \widehat{\mathbf{x}}_t \widehat{\mathbf{x}}_t^{\mathsf{T}} L_t^{\mathsf{T}} C_i^{\mathsf{T}},$$

where $\Sigma_t = \text{Cov}[\mathbf{x}_t | \widehat{\mathbf{x}}_t]$ is the conditional estimation error covariance (initialized from Σ_1, which is given). When the control-dependent, state-dependent, and internal noise terms are removed ($C_1, \ldots, C_c = 0$, $D_1, \ldots, D_d = 0$, $\Omega^\eta = 0$), equation 5.3 reduces to the Kalman filter in equation 3.2. Note that using equation 5.3 instead of equation 5.2 online reduces the total expected cost because equation 5.3 achieves lower estimation error than any other linear filter, and the expected cost depends on the conditional estimation error covariance. This can be seen from

$$\text{E}[v_t(\mathbf{x}_t, \widehat{\mathbf{x}}_t) | \widehat{\mathbf{x}}_t] = \widehat{\mathbf{x}}_t^{\mathsf{T}} S_t^x \widehat{\mathbf{x}}_t + s_t + \text{tr}\big((S_t^x + S_t^e)\text{Cov}[\mathbf{x}_t | \widehat{\mathbf{x}}_t]\big)$$

6 Application to Reaching Movements

We now illustrate how the methodology developed above can be used to construct models relevant to motor control. Since this is a methodological rather than a modeling article, a detailed evaluation of the resulting models in the context of the motor control literature will not be given here. The first model is a one-dimensional model of reaching, and includes control-dependent noise but no state-dependent or internal noise. The latter two forms of noise are illustrated in the second model, where we estimate the position of a stationary peripheral target without making a movement.

6.1 Models. We model a single-joint movement (such as flexing the elbow) that brings the hand to a specified target. For simplicity, the rotational motion is replaced with translational motion; the hand is modeled as a point mass ($m = 1$ kg) whose one-dimensional position at time t is $p(t)$. The combined action of all muscles is represented with the force $f(t)$ acting on the hand. The control signal $u(t)$ is transformed into force $f(t)$ by adding control-dependent multiplicative noise and applying a second-order muscle-like low-pass filter (Winter, 1990) of the form $\tau_1\tau_2 \ddot{f}(t) + (\tau_1 + \tau_2)\dot{f}(t) + f(t) = u(t)$, with time constants $\tau_1 = \tau_2 = 0.04$ sec. Note that a second-order filter can be written as a pair of coupled first-order filters (with outputs g and f) as follows: $\tau_1\dot{g}(t) + g(t) = u(t)$, $\tau_2\dot{f}(t) + f(t) = g(t)$.

The task is to move the hand from the starting position $p(0) = 0$ m to the target position $p^* = 0.1$ m and stop there at time t_{end}, with minimal energy

consumption. Movement durations are in the interval $t_{end} \in [0.25$ sec; 0.35 sec]. Time is discretized at $\Delta = 0.01$ sec. The total cost is defined as

$$(p(t_{end}) - p^*)^2 + (w_v \dot{p}(t_{end}))^2 + (w_f f(t_{end}))^2 + \frac{r}{n-1} \sum_{k=1}^{n-1} u(k\Delta)^2.$$

The first term enforces positional accuracy; the second and third terms specify that the movement has to stop at time t_{end}, that is, both the velocity and force have to vanish; and the last term penalizes energy consumption. It makes sense to set the scaling weights w_v and w_f so that $w_v \dot{p}(t)$ and $w_f f(t)$ averaged over the movement have magnitudes similar to the hand displacement $p^* - p(0)$. For a 0.1 m reaching movement that lasts about 0.3 sec, these weights are $w_v = 0.2$ and $w_f = 0.02$. The weight of the energy term was set to $r = 0.00001$.

The discrete-time system state is represented with the five-dimensional vector

$$\mathbf{x}_t = [p(t); \dot{p}(t); f(t); g(t); p^*]$$

initialized from a gaussian with mean $\widehat{\mathbf{x}}_1 = [0; 0; 0; 0; p^*]$. The auxiliary state variable $g(t)$ is needed to implement a second-order filter. The target p^* is included in the state so that we can capture the above cost function using a quadratic with no linear terms: defining $\mathbf{p} = [1; 0; 0; 0; -1]$, we have $\mathbf{p}^\mathsf{T} \mathbf{x}_t = p(t_{end}) - p^*$, and so $\mathbf{x}_t^\mathsf{T}(\mathbf{p}\mathbf{p}^\mathsf{T})\mathbf{x}_t = (p(t_{end}) - p^*)^2$. Note that the same could be accomplished by setting $\mathbf{p} = [1; 0; 0; 0; -p^*]$ and $\mathbf{x}_t = [p(t); \dot{p}(t); f(t); g(t); 1]$. The advantage of the formulation used here is that because the target is represented in the state, the same control law can be reused for other targets. The control law, of course, depends on the filter, which depends on the initial expected state, which depends on the target— and so a control law optimal for one target is not necessarily optimal for all other targets. Unpublished simulation results indicate good generalization, but a more detailed investigation of how the optimal control law depends on the target position is needed.

The sensory feedback carries information about position, velocity, and force:

$$\mathbf{y}_t = [p(t); \dot{p}(t); f(t)] + \boldsymbol{\omega}_t.$$

The vector $\boldsymbol{\omega}_t$ of sensory noise terms has zero-mean gaussian distribution with diagonal covariance,

$$\Omega^\omega = (\sigma_s \mathrm{diag}[0.02 \text{ m}; 0.2 \text{ m/s}; 1 \text{ N}])^2,$$

where the relative magnitudes are set using the same order-of-magnitude reasoning as before, and $\sigma_s = 0.5$. The multiplicative noise term added to the discrete-time control signal $u_t = u(t)$ is $\sigma_c \varepsilon_t u_t$, where $\sigma_c = 0.5$. Note that

σ_c is a unitless quantity that defines the noise magnitude relative to the control signal magnitude.

The discrete-time dynamics of the above system are

$$
\begin{aligned}
p(t + \Delta) &= p(t) + \dot{p}(t)\Delta \\
\dot{p}(t + \Delta) &= \dot{p}(t) + f(t)\Delta/m \\
f(t + \Delta) &= f(t)(1 - \Delta/\tau_2) + g(t)\Delta/\tau_2 \\
g(t + \Delta) &= g(t)(1 - \Delta/\tau_1) + u(t)(1 + \sigma_c\varepsilon_t)\Delta/\tau_1,
\end{aligned}
$$

which is transformed into the form of equation 3.1 by the matrices

$$
A = \begin{bmatrix}
1 & \Delta & 0 & 0 & 0 \\
0 & 1 & \Delta/m & 0 & 0 \\
0 & 0 & 1 - \Delta/\tau_2 & \Delta/\tau_2 & 0 \\
0 & 0 & 0 & 1 - \Delta/\tau_1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\quad
B = \begin{bmatrix}
0 \\ 0 \\ 0 \\ \Delta/\tau_1 \\ 0
\end{bmatrix}
$$

$$
H = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0
\end{bmatrix}
$$

$$
C_1 = B\sigma_c; c = 1; d = 0
$$
$$
\Sigma_1 = \Omega^\xi = \Omega^\eta = 0.
$$

The cost matrices are $R = r$, $Q_{1,\dots,n-1} = 0$, and $Q_n = \mathbf{p}\mathbf{p}^\mathsf{T} + \mathbf{v}\mathbf{v}^\mathsf{T} + \mathbf{f}\mathbf{f}^\mathsf{T}$, where

$$
\mathbf{p} = [1; 0; 0; 0; -1]; \quad \mathbf{v} = [0; w_v; 0; 0; 0]; \quad \mathbf{f} = [0; 0; w_f; 0; 0].
$$

This completes the formulation of the first model. The above algorithm can now be applied to obtain the control law and filter, and the closed-loop system can be simulated. To replace the control-dependent noise with additive noise of similar magnitude (and compare the effects of the two forms of noise), we will set $c = 0$ and $\Omega^\xi = (4.6\ \text{N})^2 B B^\mathsf{T}$. The value of 4.6 N is the average magnitude of the control-dependent noise over the range of movement durations (found through 10,000 simulation runs at each movement duration).

We also model an estimation process under state-dependent and internal noise, in the absence of movement. In that case, the state is $x_t = p^*$, where the stationary target p^* is sampled from a gaussian with mean $\widehat{x}_1 \in \{5\ \text{cm}, 15\ \text{cm}, 25\ \text{cm}\}$ and variance $\Sigma_1 = (5\ \text{cm})^2$. Note that target eccentricity is represented as distance rather than visual angle. The state-dependent noise has scale $D_1 = 0.5$, fixation is assumed to be at 0 cm, the time step is $\Delta = 10$ msec, and we run the estimation process for $n = 100$ time steps. In one set of simulations, we use internal noise $\Omega^\eta = (0.5\ \text{cm})^2$ without additive noise. In another set of simulations, we study additive noise with the same magnitude $\Omega^\omega = (0.5\ \text{cm})^2$, without internal noise. There is no actuator to be controlled, so we have $A = H = 1$ and $B = L = 0$. Estimation is based on the adaptive filter from equation 5.3.

6.2 Results. Reaching movements are known to have stereotyped bell-shaped speed profiles (Flash & Hogan, 1985). Models of this phenomenon have traditionally been formulated in terms of deterministic open-loop minimization of some cost function. Cost functions that penalize physically meaningful quantities (such as duration or energy consumption) did not agree with empirical data (Nelson, 1983); in order to obtain realistic speed profiles, it appeared necessary to minimize a smoothness-related cost that penalizes the derivative of acceleration (Flash & Hogan, 1985) or torque (Uno et al., 1989). Smoothness-related cost functions have also been used in the context of stochastic optimal feedback control (Hoff, 1992) to obtain bell-shaped speed profiles. It was recently shown, however, that smoothness does not have to be explicitly enforced by the cost function; open-loop minimization of end-point error was found sufficient to produce realistic trajectories, provided that the multiplicative nature of motor noise is taken into account (Harris & Wolpert, 1998). While this is an important step toward a more principled optimization model of trajectory smoothness, it still contains an ad hoc element: the optimization is performed in an open loop, which is suboptimal, especially for movements of longer duration. Our model differs from Harris and Wolpert (1998) in that not only the average sequence of control signals is optimal, but the feedback gains that determine the online sensory-guided adjustments are also optimal. Optimal feedback control of reaching has been studied by Meyer, Abrams, Kornblum, Wright, and Smith (1988) in an intermittent setting, and Hoff (1992) in a continuous setting. However, both of these models assume full state observation. Ours is the first optimal control model of reaching that incorporates sensory noise and combines state estimation and feedback control into an optimal sensorimotor loop. The predicted movement kinematics shown in Figure 1A closely resemble observed movement trajectories (Flash & Hogan, 1985).

Another well-known property of reaching movements, first observed a century ago by Woodworth and later quantified as Fitts' law, is the trade-off between speed and accuracy. The fact that faster movements are less accurate implies that the instantaneous noise in the motor system is control-dependent, in agreement with direct measurements of isometric force fluctuations (Sutton and Sykes, 1967; Schmidt et al., 1979; Todorov, 2002) that show standard deviation increasing linearly with the mean. Naturally, this noise scaling has formed the basis of both closed-loop (Meyer et al., 1988; Hoff, 1992) and open-loop (Harris & Wolpert, 1998) optimization models of the speed-accuracy trade-off. Figure 1B illustrates the effect in our model: as the (specified) movement duration increases, the standard deviation of the end-point error achieved by the optimal controller decreases. To emphasize the need for incorporating control-dependent noise, we modified the model by making the noise in the plant dynamics additive, with fixed magnitude chosen to match the average multiplicative noise magnitude over the range of movement durations. With that change, the end-point error showed the opposite trend to the one observed experimentally (see Figure 1B).
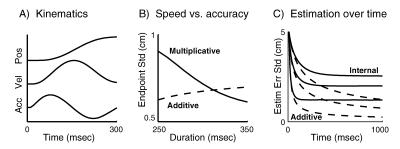
Figure 1: (A) Normalized position (Pos), velocity (Vel), and acceleration (Acc) of the average trajectory of the optimal controller. (B) A separate optimal controller was constructed for each instructed duration, the resulting closed-loop system was simulated for 10,000 trials, and the positional standard deviation at the end of the trial was plotted. This was done with either **multiplicative** (solid line) or **additive** (dashed line) noise in the plant dynamics. (C) The position of a stationary peripheral target was estimated over time, under **internal** estimation noise (solid line) or **additive** observation noise (dashed line). This was done in three sets of trials, with target positions sampled from gaussians with means 5 cm (bottom), 15 cm (middle), and 25 cm (top). Each curve is an average over 10,000 simulation runs.

It is interesting to compare the effects of the control penalty r and the multiplicative noise scaling σ_c. As equation 4.2 shows, both terms penalize large control signals—directly in the case of r and indirectly (via increased uncertainty) in the case of σ_c. Consequently, both terms lead to a negative bias in end-point position (not shown), but the effect is much more pronounced for r. Another consequence of the fact that larger controls are more costly arises in the control of redundant systems, where the optimal strategy is to follow a minimal intervention principle, that is, to leave task-irrelevant deviations from the average behavior uncorrected (Todorov & Jordan, 2002a, 2002b). Simulations have shown that this more complex effect is dependent on σ_c and actually decreases when r is increased while σ_c is kept constant (Todorov & Jordan, 2002b).

Figure 1C shows simulation results from our second model, where the position of a stationary peripheral target is estimated by the optimal adaptive filter in equation 5.3, operating under internal estimation noise or additive observation noise of the same magnitude. In each case, we show results for three sets of targets with varying average eccentricity. The standard deviations of the estimation error always reach an asymptote (much faster in the case of internal noise). In the presence of internal noise, this asymptote depends on target eccentricity; for the chosen model parameters, the dependence is in quantitative agreement with our experimental results (Todorov, 1998). Under additive noise, the error always asymptotes to 0.
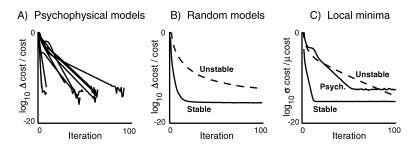
Figure 2: Relative change in expected cost as a function of iteration number, in (A) psychophysical models and (B) random models. (C) Relative variability (SD/mean) among expected costs obtained from 100 different runs of the algorithm on the same model (average over models in each class).

7 Convergence Properties

We studied the convergence properties of the algorithm in 10 models of psychophysical experiments taken from Todorov and Jordan (2002b) and 200 randomly generated models. The psychophysical models had dynamics and cost functions similar to the above example. They included two models of planar reaching, three models of passing through sequences of targets, one model of isometric force production, three models of tracking and reaching with a mechanically redundant arm, and one model of throwing. The dimensionalities of the state, control, and feedback were between 5 and 20, and the horizon n was about 100. The psychophysical models included control-dependent dynamics noise and additive observation noise, but no internal or state-dependent noise. The details of all these models are interesting from a motor control point of view, but we omit them here since they did not affect the convergence of the algorithm in any systematic way.

The random models were divided into two groups of 100 each: passively stable, with all eigenvalues of A being smaller than 1, and passively unstable, with the largest eigenvalue of A being between 1 and 2. The dynamics were restricted so that the last component of $\mathbf{x}_t$ was 1—to make the random models more similar to the psychophysical models, which always incorporated a constant in the state description. The state, control, and measurement dimensionalities were sampled uniformly between 5 and 20. The random models included all forms of noise allowed by equation 3.1.

For each model, we initialized $K_{1,...,n-1}$ from equation 3.2 and applied our iterative algorithm. In all cases convergence was very rapid (see Figures 2A and 2B), with the relative change in expected cost decreasing exponentially. The jitter observed at the end of the minimization (see Figure 2A) is due to numerical round-off errors (note the log scale) and continues indefinitely. The exponential convergence regime does not always start from the first iteration (see Figure 2A). Similar behavior was observed for the absolute

change in expected cost (not shown). As one would expect, random models with unstable passive dynamics converged more slowly than passively stable models. Convergence was observed in all cases.

To test for the existence of local minima, we focused on five psychophysical, five random stable, and five random unstable models. For each model, the algorithm was initialized 100 times with different randomly chosen sequences $K_{1,...,n-1}$, and run for 100 iterations. For each model, we computed the standard deviation of the expected cost obtained at each iteration and divided by the mean expected cost at that iteration. The results, averaged within each model class, are plotted in Figure 2C. The negligibly small values after convergence indicate that the algorithm always finds the same solution. This was true for every model we studied, despite the fact that the random initialization sometimes produced very large initial costs. We also examined the K and L sequences found at the end of each run, and the differences seemed to be due to round-off errors. Thus, we conjecture that the algorithm always converges to the globally optimal solution. So far we have not been able to prove this analytically and cannot offer a satisfying intuitive explanation at this time.

Note that the system can be unstable even for the optimal controller. Formally, that does not affect the derivation, because in a discrete-time finite-horizon system, all numbers remain finite. In practice, the components of $\mathbf{x}_t$ can exceed the maximum floating-point number whenever the eigenvalues of $(A - BL_t)$ are sufficiently large. In the applications we are interested in (Todorov, 1998; Todorov & Jordan, 2002b), such problems were never encountered.

8 Improving Performance via Adaptive Estimation

Although the iterative algorithm given by equations 4.2 and 5.2 is guaranteed to converge, and empirically it appears to converge to the globally optimal solution, performance can still be suboptimal due to the imposed restriction to nonadaptive filters. Here we present simulations aimed at quantifying this suboptimality.

Because the potential suboptimality arises from the restriction to nonadaptive filters, it is natural to ask what would happen if that restriction were removed in run time and the optimal adaptive linear filter from equation 5.3 were used instead. Recall that although the control law is optimized under the assumption of a nonadaptive filter, it yields better performance if a different filter, which somehow achieves lower estimation error, is used in run time. Thus, in our first test, we simply replace the nonadaptive filter with equation 5.3 in run time and compute the reduction in expected total cost.

The above discussion suggests a possibility for further improvement. The control law is optimal with respect to some sequence of filter gains $K_{1,...,n-1}$. But the adaptive filter applied in run time uses systematically different gains, because it achieves systematically lower estimation error. We can run

Table 2: Cost Reduction.

	Model		
Method	Psychophysical	Random Stable	Random Unstable
Adaptive Estimator	1.9 %	0 %	31.4 %
Reoptimized Controller	1.7	0	28.3

Notes: Numbers indicate percent reduction in expected total cost, relative to the cost of the solution found by our iterative algorithm. The two improvement methods are described in the text. Each method is applied to 10 models in each model class. For each model and method, expected total cost is computed from 10,000 simulation runs. A value of 0% indicates that with a sample size of 10 models, the improvement was not significantly different from 0% (t-test, $p = 0.05$ threshold).

our control law in conjunction with the adaptive filter and find the average filter gains $\widetilde{K}_{1,\dots,n-1}$ that are used online. Now, one would think that if we reoptimized the control law for the nonadaptive filter $\widetilde{K}_{1,\dots,n-1}$, which better reflects the gains being used online by the adaptive filter, this will further improve performance. This is the second test we apply.

As Table 2 shows, neither method improves performance substantially for psychophysical models or random stable models. However, both methods result in substantial improvement for random unstable models. This is not surprising. In the passively stable models, the differences between the expected and actual values of the states and controls are relatively small, and so the optimal nonadaptive filter is not that different from the optimal adaptive filter. The unstable models, on the other hand, are very sensitive to small perturbations and thus follow substantially different state-control trajectories in different simulation runs. So the advantage of adaptive filtering is much greater. Since musculoskeletal plants have stable passive dynamics, we conclude that our algorithm is well suited for approximating the optimal sensorimotor system.

It is interesting that control law reoptimization in addition to adaptive filtering is actually worse than adaptive filtering alone—contrary to our intuition. This was the case for every model we studied. Although it is not clear where the problem with the reoptimization method lies, this somewhat unexpected result provides further justification for the restriction we introduced. In particular, it suggests that the control law that is optimal under the best nonadaptive filter may be close to optimal under the best adaptive filter.

9 Discussion

We have presented an algorithm for stochastic optimal control and estimation of partially observable linear dynamical systems, subject to quadratic costs and noise processes characteristic of the sensorimotor system (see

equation 3.1). We restricted our attention to controllers that use state estimates obtained by nonadaptive linear filters. The optimal control law for any such filter was shown to be linear, as given by equation 4.2. The optimal nonadaptive linear filter for any linear control law is given by equation 5.2. Iteration of equations 4.2 and 5.2 is guaranteed to converge to a filter and a control law optimal with respect to each other. We found numerically that convergence is exponential, local minima do not to exist, and the effects of assuming nonadaptive filtering are negligible for the control problems of interest. The application of the algorithm was illustrated in the context of reaching movements. The optimal adaptive filter, equation 5.3, as well as the optimal controller for the fully observable case, equation 4.3, were also derived. To facilitate the application of our algorithm in the field of motor control and elsewhere, we have made a Matlab implementation available at www.cogsci.ucsd.edu/~todorov.

While our work was motivated by models of biological movement, the results presented here could be of interest to a wider audience. Problems with multiplicative noise have been studied in the optimal control literature, but most of that work has focused on the fully observable case (Kleinman, 1969; McLane, 1971; Willems & Willems, 1976; Bensoussan, 1992; El Ghaoui, 1995; Beghi & D'Alessandro, 1998; Rami et al., 2001). Our equation 4.3 is consistent with these results. The partially observable case that we addressed (and that is most relevant to models of sensorimotor control) is much more complex, because the independence of estimation and control breaks down in the presence of signal-dependent noise. The work most similar to ours is Pakshin (1978) for discrete-time dynamics and Phillis (1985) for continuous-time dynamics. These authors addressed a closely related problem using a different methodology. Instead of analyzing the closed-loop system directly, the filter and control gains were treated as open-loop controls to a modified deterministic dynamical system, whose cost function matches the expected cost of the original system. With that transformation, it is possible to use Pontryagin's maximum principle, which is applicable only to deterministic open-loop control, and obtain necessary conditions that the optimal filter and control gains must satisfy. Although our results were obtained independently, we have been able to verify that they are consistent with Pakshin (1978) by removing from our model the internal estimation noise (which to our knowledge has not been studied before); combining equations 4.2 and 5.2; and applying certain algebraic transformations. However, our approach has three important advantages. First, we managed to prove that the optimal control law is linear under a nonadaptive filter, while this linearity had to be assumed before. Second, using the optimal cost-to-go function to derive the optimal filter revealed that adaptive filtering improves performance, even though the control law is optimized for a nonadaptive filter. And most important, our approach yields a coordinate-descent algorithm with guaranteed convergence, as well as appealing numerical properties illustrated in sections 7 and 8. Each of the two steps of our

coordinate-descent algorithm is computed efficiently in a single pass through time. In contrast, application of Pontryagin's maximum principle yields a system of coupled difference (Pakshin, 1978) or differential (Phillis, 1985) equations with boundary conditions at the initial and final time, but no algorithm for solving that system. In other words, earlier approaches obscure the key property we uncovered: that half of the problem can be solved efficiently given a solution to the other half.

Finally, there may be an efficient way to obtain a control law that achieves better performance under adaptive filtering. Our attempt to do so through reoptimization (see section 8) failed, but another approach is possible. Using the optimal adaptive filter (see equation 5.3) would make $E[v_{t+1}]$ a complex function of $\widehat{\mathbf{x}}_t$, $\mathbf{u}_t$, and the resulting v_t would no longer be in the assumed parametric form (which is why we introduced the restriction to nonadaptive filters). But we could force that complex v_t in the desired form by approximating it with a quadratic in $\widehat{\mathbf{x}}_t$, $\mathbf{u}_t$. This yields additional terms in equation 4.2. We have pursued this idea in our earlier work (Todorov, 1998); an independent but related method has been developed by Moore, Zhou, and Lim (1999). The problem with such approximations is that convergence guarantees no longer seem possible. While Moore et al. did not illustrate their method with numerical examples, in our work we have found that the resulting algorithm is not always stable. These difficulties convinced us to abandon the earlier idea in favor of the methodology presented here. Nevertheless, approximations that take adaptive filtering into account may yield better control laws under certain conditions and deserve further investigation. Note, however, that the resulting control laws will have to be used in conjunction with an adaptive filter, which is much less efficient in terms of online computation.

References

Anderson, F., & Pandy, M. (2001). Dynamic optimization of human walking. *J Biomech. Eng*, *123*(5), 381–390.

Beghi, A., & D'Alessandro, D. (1998). Discrete-time optimal control with control-dependent noise and generalized Riccati difference equations. *Automatica*, *34*, 1031–1034.

Bensoussan, A. (1992). *Stochastic control of partially observable systems*. Cambridge: Cambridge University Press.

Bertsekas, D., & Tsitsiklis, J. (1997). *Neuro-dynamic programming*. Belmont, MA: Athena Scientific.

Burbeck, C., & Yap, Y. (1990). Two mechanisms for localization? Evidence for separation-dependent and separation-independent processing of position information. *Vision Research*, *30*(5), 739–750.

Chow, C., & Jacobson, D. (1971). Studies of human locomotion via optimal programming. *Math Biosciences*, *10*, 239–306.

Davis, M., & Vinter, R. (1985). *Stochastic modelling and control*. London: Chapman and Hall.

El Ghaoui, L. (1995). State-feedback control of systems of multiplicative noise via linear matrix inequalities. *Systems and Control Letters*, *24*, 223–228.

Flash, T., & Hogan, N. (1985). The coordination of arm movements: An experimentally confirmed mathematical model. *Journal of Neuroscience*, *5*(7), 1688–1703.

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

Hatze, H., & Buys, J. (1977). Energy-optimal controls in the mammalian neuromuscular system. *Biol. Cybern.*, *27*(1), 9–20.

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

Jacobson, D., & Mayne, D. (1970). *Differential dynamic programming*. New York: Elsevier.

Jones, K., Hamilton, A., & Wolpert, D. (2002). Sources of signal-dependent noise during isometric force production. *Journal of Neurophysiology*, *88*, 1533–1544.

Kleinman, D. (1969). Optimal stationary control of linear systems with control-dependent noise. *IEEE Transactions on Automatic Control*, *AC-14*(6), 673–677.

Kording, K., & Wolpert, D. (2004). The loss function of sensorimotor learning. *Proceedings of the National Academy of Sciences*, *101*, 9839–9842.

Kuo, A. (1995). An optimal control model for analyzing human postural balance. *IEEE Transactions on Biomedical Engineering*, *42*, 87–101.

Kushner, H., & Dupuis, P. (2001). *Numerical methods for stochastic optimal control problems in continuous time* (2nd ed.). New York: Springer.

Li, W., & Todorov, E. (2004). Iterative linear-quadratic regulator design for nonlinear biological movement systems. In *First International Conference on Informatics in Control, Automation and Robotics*, vol. 1, 222–229. N.P.: INSTICC Press.

Loeb, G., Levine, W., & He, J. (1990). Understanding sensorimotor feedback through optimal control. *Cold Spring Harb. Symp. Quant. Biol.*, *55*, 791–803.

McLane, P. (1971). Optimal stochastic control of linear systems with state- and control-dependent disturbances. *IEEE Transactions on Automatic Control*, *AC-16*(6), 793–798.

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

Moore, J., Zhou, X., & Lim, A. (1999). Discrete time LQG controls with control dependent noise. *Systems and Control Letters*, *36*, 199–206.

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

Pakshin, P. (1978). State estimation and control synthesis for discrete linear systems with additive and multiplicative noise. *Avtomatika i Telemetrika*, *4*, 75–85.

Phillis, Y. (1985). Controller design of systems with multiplicative noise. *IEEE Transactions on Automatic Control*, *AC-30*(10), 1017–1019.

Rami, M., Chen, X., & Moore, J. (2001). Solvability and asymptotic behavior of generalized Riccati equations arising in indefinite stochastic LQ problems. *IEEE Transactions on Automatic Control*, *46*(3), 428–440.

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

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

Sutton, R., & Barto, A. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Todorov, E. (1998). *Studies of goal-directed movements*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Todorov, E. (2002). Cosine tuning minimizes motor errors. *Neural Computation*, *14*(6), 1233–1260.

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature Neuroscience*, *7*(9), 907–915.

Todorov, E., & Jordan, M. (2002a). A minimal intervention principle for coordinated movement. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems, 15* (pp. 27–34). Cambridge, MA: MIT Press.

Todorov, E., & Jordan, M. (2002b). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, *5*(11), 1226–1235.

Todorov, E., & Li, W. (2004). *A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems*. Manuscript submitted for publication.

Uno, Y., Kawato, M., & Suzuki, R. (1989). Formation and control of optimal trajectory in human multijoint arm movement: Minimum torque-change model. *Biological Cybernetics*, *61*, 89–101.

Whitaker, D., & Latham, K. (1997). Disentangling the role of spatial scale, separation and eccentricity in Weber's law for position. *Vision Research*, *37*(5), 515–524.

Whittle, P. (1990). *Risk-sensitive optimal control*. New York: Wiley.

Willems, J. L., & Willems, J. C. (1976). Feedback stabilizability for stochastic systems with state and control dependent noise. *Automatica*, *1976*, 277–283.

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