

Supplementary Material

Learning optimal adaptation strategies in unpredictable motor tasks

D.A. Braun, A. Aertsen, D.M. Wolpert, C. Mehring

Contents

1	Adaptive Optimal Control Methods	2
1.1	The Estimation Problem	3
1.2	The Control Problem	4
1.3	Connection to Non-adaptive Optimal Control Models	5
1.4	Arm Model	6
2	Model Fit	8

1 Adaptive Optimal Control Methods

The general mathematical model underlying the fits and predictions in the main text belongs to a class of modified Linear-Quadratic-Gaussian (LQG) models [Stengel, 1994]. LQG models deal with linear dynamic systems, quadratic cost functions as performance criteria, and Gaussian random variables as noise. Here we consider the following model class:

$$\begin{aligned}
 \vec{x}_{t+1} &= F[\vec{a}] \vec{x}_t + G \vec{u}_t + \vec{\xi}_t + G \sum_i C_i \vec{u}_t \sigma_{i,t} \\
 \vec{y}_t &= H \vec{x}_t + \vec{\chi}_t \\
 J &= \frac{1}{2} \mathbb{E} \left[\sum_{t=0}^{\infty} \left\{ \vec{x}_t^T Q \vec{x}_t + \vec{u}_t^T R \vec{u}_t \right\} \right]
 \end{aligned} \tag{1}$$

with the following variables

dynamic state	$\vec{x}_t \in \mathfrak{R}^n$
unknown system parameters	$\vec{a} \in \mathfrak{R}^l$
control signal	$\vec{u}_t \in \mathfrak{R}^m$
feedback observation	$\vec{y}_t \in \mathfrak{R}^q$
expected cumulative cost	$J \in \mathfrak{R}$
state cost matrix	$Q = Q^T \geq 0$
control cost matrix	$R = R^T > 0$

Time is discretized in bins of $10ms$. The noise variables $\vec{\xi}_t \in \mathfrak{R}^n$, $\vec{\chi}_t \in \mathfrak{R}^k$, $\sigma_{i,t} \in \mathfrak{R}$ are realizations of independent, zero-mean, Gaussian noise processes with covariance matrices $\mathbb{E}[\vec{\xi}_{t_1} \vec{\xi}_{t_2}^T] = \Omega_\xi \delta_{t_1 t_2}$, $\mathbb{E}[\vec{\chi}_{t_1} \vec{\chi}_{t_2}^T] = \Omega_\chi \delta_{t_1 t_2}$ and $\mathbb{E}[\sigma_{i_1, t_1} \sigma_{i_2, t_2}] = \delta_{t_1 t_2} \delta_{i_1 i_2}$ respectively. The dynamic state \vec{x}_t is a hidden variable that needs to be inferred from feedback observations \vec{y}_t . An initial estimate of \vec{x}_0 is given by a normal distribution with mean \vec{x}_0^x and covariance P_0^x . Accordingly, an initial estimate of the unknown parameters \vec{a} is given by a normal distribution with mean \vec{a}_0^a and covariance P_0^a . This allows to state the optimal control problem: given $F[\vec{a}_0]$, G , H , C_i , Ω_ξ , Ω_χ , P_0^x , P_0^a , R , Q , what is the control law $\vec{u}_t = \vec{\pi}(\vec{x}_0, \vec{u}_0, \dots, \vec{u}_{t-1}, \vec{y}_0, \dots, \vec{y}_{t-1}, t)$ that minimizes the expected cumulative cost J ?

In the absence of multiplicative noise (i.e. $C_i \equiv 0 \forall i$) and assuming perfect knowledge of all system parameters \vec{a} , the posed optimal control problem has a well-known solution [Stengel, 1994]: a Kalman filter estimates the hidden state \vec{x}_t optimally in a least-squares sense and a linear optimal controller maps this estimate \vec{x}_t into a control signal \vec{u}_t . Several approximative solutions have been suggested in the literature to solve the non-adaptive control problem with multiplicative noise [Moore et al., 1999; Todorov, 2005]. Here we address the optimal control problem with multiplicative noise in the presence of parameter uncertainties.

Unfortunately, adaptive optimal control problems can, in general, neither be solved analytically nor numerically. Therefore, reasonable approximations have to be found that are applicable to broad

classes of problems. In movement neuroscience, usually ‘indirect’ adaptive control schemes are used, implying that subjects avail themselves of internal models both to predict their environment and to adjust their motor control on the basis of these predictions. Mathematically, this entails the separation of estimation and control processes, i.e. the general proceeding is (1) to identify the system parameters \vec{a} on-line, and (2) to exploit the resulting estimate $\vec{\hat{a}}_t$ by appropriately adjusting the control law $\vec{\pi}$ when computing \vec{u}_t .

1.1 The Estimation Problem

To perform system identification on-line in a noisy environment implies solving a *joint filtering problem* [Haykin, 2001], because states and parameters have to be estimated simultaneously. Joint filtering methods are based on the definition of an *augmented* or *joint* state space with the concatenated state vector

$$\vec{\mathbf{x}}_t = \begin{bmatrix} \vec{x}_t \\ \vec{a}_t \end{bmatrix} \quad (2)$$

Since the unknown parameters are assumed to be constant ($\vec{a}_{t+1} = \vec{a}_t$), system identification can be simply instantiated by letting the parameters do a random walk driven by a process noise¹ $\vec{v}_t \sim \mathcal{N}(0, \Omega_\nu)$

$$\vec{a}_{t+1} = \vec{a}_t + \vec{v}_t \quad (3)$$

To be compatible with the concatenated state vector, the state transition matrix, the measurement matrix and the process covariance matrix need to be modified for the joint state space

$$\begin{aligned} \tilde{F}[\vec{a}_t] &= \begin{bmatrix} F[\vec{a}_t] & 0 \\ 0 & \mathbb{I}_{l \times l} \cdot \vec{a}_t \end{bmatrix} \\ \tilde{H} &= \begin{bmatrix} H & 0 \end{bmatrix} \\ \tilde{\Omega}_{\tilde{\xi}} &= \begin{bmatrix} \Omega_\xi & 0 \\ 0 & \Omega_\nu \end{bmatrix} \end{aligned}$$

Since adaptive control problems are inherently nonlinear, the standard Kalman filter solution [Kalman, 1960] is not applicable in the augmented state space. A state-of-the-art method for nonlinear estimation problems is *Unscented Kalman filtering* [Haykin, 2001], where the distribution of the random variable is sampled efficiently by carefully chosen *sigma points* that are propagated through the full nonlinearity. The *sigma vectors* of the random variable with mean $\hat{\vec{\mathbf{x}}} \in \mathfrak{R}^{n+l}$ and covariance P^ξ are calculated according to

$$\begin{aligned} \mathcal{X}_0 &= \hat{\vec{\mathbf{x}}} \\ \mathcal{X}_i &= \hat{\vec{\mathbf{x}}} + \gamma(\sqrt{P^\xi})_i & i = 1, \dots, n+l \\ \mathcal{X}_i &= \hat{\vec{\mathbf{x}}} - \gamma(\sqrt{P^\xi})_{i-n-l} & i = n+l+1, \dots, 2(n+l) \end{aligned}$$

¹The parameter covariance matrix Ω_ν determines the time scale of parameter adaptation. In the present case, Ω_ν is a phenomenological constant that captures the adaptation rate of the brain for a specific learning task. The optimal control problem is posed under the constraint of this given adaptation rate.

with the scaling parameter γ [Julier et al., 1995]. The expression $(\sqrt{P^x})_i$ denotes the i th column of the matrix square root of P^x that can be determined, for instance, by the lower-triangular Cholesky factorization. This leads to the following Kalman filter equations:

$$\vec{\mathbf{x}}_t = \vec{\mathbf{x}}_t^- + K_t [\vec{y}_t - \vec{y}_t^-] \quad (4)$$

$$P_t^x = P_t^{x-} - K_t P_t^{yy} K_t^T \quad (5)$$

with the Kalman gain $K_t = P_t^{xy} (P_t^{yy})^{-1}$ and the covariances

$$P_t^{yy} = \sum_{i=0}^{2n} W_i^{(c)} (\mathcal{Y}_t^- - \vec{y}_t^-) (\mathcal{Y}_t^- - \vec{y}_t^-)^T + \Omega_\chi \quad (6)$$

$$P_t^{xy} = \sum_{i=0}^{2n} W_i^{(c)} (\mathcal{X}_t^- - \vec{\mathbf{x}}_t^-) (\mathcal{Y}_t^- - \vec{y}_t^-)^T \quad (7)$$

$$P_t^{x-} = \sum_{i=0}^{2n} W_i^{(c)} (\mathcal{X}_t^- - \vec{\mathbf{x}}_t^-) (\mathcal{X}_t^- - \vec{\mathbf{x}}_t^-)^T + \tilde{\Omega}_\xi + \sum_i \tilde{G} C_i \vec{u}_t \vec{u}_t^T C_i^T \tilde{G}^T \quad (8)$$

The last summand of equation (8) accounts for higher variability due to multiplicative noise and is derived from a linear approximation scheme following [Moore et al., 1999]. The required sigma points are calculated as

$$\mathcal{X}_t^- = \tilde{F}[\mathcal{X}_{t-1}] + \tilde{G} \vec{u}_{t-1} \quad (9)$$

$$\vec{\mathbf{x}}_t^- = \sum_{i=1}^{2n} W_i^{(m)} \mathcal{X}_t^- \quad (10)$$

$$\mathcal{Y}_t^- = \tilde{H} \mathcal{X}_t^- \quad (11)$$

$$\vec{y}_t^- = \sum_{i=1}^{2n} W_i^{(m)} \mathcal{Y}_t^- \quad (12)$$

with scaling parameters $W_i^{(m)}$ and $W_i^{(c)}$ [Julier et al., 1995].

1.2 The Control Problem

In general, the posed adaptive optimal control problem will be a *dual control problem*² without a straightforward solution [Åström and Wittenmark, 1989]. In the case of partial system observability, it is a common approximation [Bar-Shalom and Tse, 1974; Bar-Shalom, 1981] to decompose the cost function J into a deterministic part J_D (*certainty-equivalent control*), a cautious part J_C (*prudent control*), and a probing part J_P (*explorative control*). Accordingly, the adaptive optimal controller

²The dual control problem is conceptually related to the *exploration-exploitation-dilemma* known in reinforcement learning [Sutton and Barto, 1998], since it deals with a similar set of questions: If the parameter uncertainty is not too high, should one act as if there were no uncertainty (*certainty-equivalent control*)? Should one be particularly prudent in an unknown environment (*cautious control*)? Or is it best to be explorative, i.e. invest short-term effort into identifying the unknown parameters and exploit this knowledge subsequently (*probing control*)?

u is designed as a composition of the sub-controllers u^D , u^C and u^P . Then it depends on the characteristics of the specific control problem which sub-controllers dominate and which might be neglected. Especially the design of the sub-controllers u^C and u^P usually follows mere heuristic principles. The design of u^D can be obtained by virtue of the *certainty-equivalence principle*³. In the present case, the following certainty-equivalent controller can be derived by applying again the approximation scheme of [Moore et al., 1999] to linearize the multiplicative noise terms

$$\vec{u}_t^D = -L_t[\vec{a}_t] \vec{x}_t \quad (13)$$

with

$$L_t[\vec{a}_t] = (R + G^T S_t G + \sum_i C_i^T G^T S_t G C_i)^{-1} G^T S_t F[\vec{a}_t] \quad (14)$$

The matrix S_t can be easily computed by solving the pertinent Riccati equation by means of established standard methods

$$S_t = Q + F[\vec{a}_t]^T S_t F[\vec{a}_t] - F[\vec{a}_t]^T S_t G (R + G^T S_t G + \sum_i C_i^T G^T S_t G C_i)^{-1} G^T S_t F[\vec{a}_t] \quad (15)$$

In case of perfect knowledge of system parameters and full state observability (i.e. $\vec{y}_t = \vec{x}_t$), the above solution can be shown analytically to be optimal [Kleinman, 1969]. In case of partial observability, equations (13)-(15) can only be part of an approximative solution [Moore et al., 1999; Todorov, 2005]. In case of parameter uncertainties, the additional difficulty arises that successful system identification in the closed loop cannot be guaranteed generically [Kumar, 1983, 1990; van Schuppen, 1994; Campi and Kumar, 1996; Campi, 1997]. Here, we only considered unknown system parameters in the state transition matrix, but the algorithm is also applicable in the face of general parameter uncertainties provided that questions of stability and closed-loop identification are clarified on a case-to-case basis. These difficulties are omnipresent in adaptive control, simply due to the immense complexity of the topic. Indeed, the vast majority of practical applications in the field that have proven to be very robust lack a thorough mathematical treatment and convergence proof [Åström and Wittenmark, 1989; Sastry and Bodson, 1989]. Here, we compare the performance of the proposed algorithm with other non-adaptive control algorithms considering a multiplicative noise structure (**Fig. S1**).

1.3 Connection to Non-adaptive Optimal Control Models

From a theoretical point of view it seems desirable to design a unified control scheme, where “learning control” equals “standard control” in the absence of parameter uncertainties, and “learning” converges to “standard” over time. The proposed approach in sections (1.1) and (1.2) fulfils this

³When neglecting u^C and u^P , the certainty-equivalence principle leads to the control scheme of the *self-tuning regulator* [Åström and Wittenmark, 1989], i.e. the current parameter estimate \vec{a}_t is employed for control as if it were the true parameter \vec{a} , while the uncertainty P_t^a of the estimate is ignored for control purposes

criterion and is, therefore, consistent. However, the question arises in how far our “standard control” corresponds to non-adaptive optimal control schemes in the literature [Todorov and Jordan, 2002].

In contrast to previous non-adaptive optimal control schemes, we have postulated optimality not for an action sequence on a predefined time interval T , but for an indefinite runtime. In the literature this is known as *infinite horizon control* as opposed to *finite horizon control* with a predefined time window T [Stengel, 1994]. We have chosen this approach, because the finite horizon setting does not allow the implementation of adaptivity in a straightforward manner⁴. Additionally, a noisy infinite horizon model naturally predicts variable movement durations, while variable movement times T in a finite horizon model have to be introduced by deliberately drawing T from a Gaussian distribution. Remarkably, the proposed control architecture is able to reproduce the speed-accuracy trade-off in the presence of multiplicative noise and can account for the speed-target distance relationship as found experimentally (cf. **Fig. S2**). However, it remains to be tested in how far actual motor behavior can be accounted for by time-independent optimal policies, and whether and in which contexts time-dependent policies are indispensable. A straightforward generalization of the present algorithm would be to allow for state-dependent feedback gains (see [Jazwinsky, 1970] for state-dependent Riccati equation (SDRE) control).

1.4 Arm Model

In the experiment described in the main text, human subjects steered a cursor on a screen to designated targets. Since the hand movement in the experiment was very confined in space ($8cm$), the hand/cursor system is modeled with linear dynamic equations. Following previous studies [Todorov, 2005; Winter, 1990] the hand is modeled as a point mass m with two-dimensional position $\vec{p}^H(t)$ and velocity $\vec{v}^H(t) = \dot{\vec{p}}^H(t)$. The combined action of all muscles on the hand is represented by the force vector $\vec{f}(t)$. The neural control signal $\vec{u}(t)$ is transformed to this force through a second-order muscle-like low-pass filter with time constants τ_1 and τ_2 . In every instant of time, the

⁴In the finite horizon setting [Harris and Wolpert, 1998; Todorov and Jordan, 2002] the argument goes that during movement execution there are no explicit constraints apart from avoiding excessive control signals, and only when the target is reached accuracy becomes an issue, i.e. in mathematical terms the cost matrix Q is zero during the movement and takes the value $Q = Q_f$ at the end of the movement. To solve this finite-horizon optimal control problem, the constraint Q_f has to be propagated back through time via the Riccati recursion determining the optimal feedback gain L_t at each point in time. Thus, target-related accuracy requirements (“minimum end-point variance” [Harris and Wolpert, 1998]) shape the optimal trajectory. Obviously, this procedure cannot be carried over to the adaptive case in a straightforward manner, since the Riccati recursion presupposes knowledge of the true system parameters to determine the entire sequence of optimal feedback gains and to backpropagate terminal constraints through time. Finally, going one step back and trying to solve the pertinent Bellman equation under parameter uncertainties is also not an option due to mathematical intractability. In contrast, a stationary feedback controller in an infinite horizon setting easily carries over to the adaptive case by re-computing the “stationary” control law in each time step, thus, considering the most recent parameter estimate. This also implies that such an adaptive controller is applicable on-line.

hand motion is mapped to a cursor motion on a screen by use of a manipulandum. This mapping can either be straightforward, or a rotation ϕ between hand movement and cursor movement can be introduced. Neglecting the dynamics of the frictionless manipulandum, the cursor position $\vec{p}(t)$ is connected to the hand position via a simple rotation operator \mathcal{D}_ϕ , i.e. $\vec{p}(t) = \mathcal{D}_\phi \vec{p}^H(t)$. Put together, this yields the following system equations

$$\vec{\ddot{p}}(t) = \frac{1}{m} \mathcal{D}_\phi \vec{f}(t) \quad (16)$$

$$\tau_1 \tau_2 \vec{\ddot{f}}(t) + (\tau_1 + \tau_2) \vec{\dot{f}}(t) + \vec{f}(t) = \vec{u}(t) \quad (17)$$

Equation (17) can be written equivalently as a pair of coupled first-order filters with outputs g and f . This allows to formulate the state space vector $\vec{x} \in \mathbb{R}^{10}$ as

$$\vec{x}(t) = \left[p^x(t) \quad v^x(t) \quad f^x(t) \quad g^x(t) \quad p_x^{\text{TARGET}} \quad p^y(t) \quad v^y(t) \quad f^y(t) \quad g^y(t) \quad p_y^{\text{TARGET}} \right]^T$$

where the target location is absorbed in the state vector. When discretizing the above equations with time bin Δ the following system matrices are obtained

$$F[\phi] = \begin{pmatrix} 1 & \Delta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & \frac{\Delta}{m} \cos(\phi) & 0 & 0 & 0 & 0 & \frac{\Delta}{m} \sin(\phi) & 0 & 0 \\ 0 & 0 & 1 - \frac{\Delta}{\tau_2} & \frac{\Delta}{\tau_2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 - \frac{\Delta}{\tau_1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \Delta & 0 & 0 & 0 \\ 0 & 0 & -\frac{\Delta}{m} \sin(\phi) & 0 & 0 & 0 & 1 & \frac{\Delta}{m} \cos(\phi) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 - \frac{\Delta}{\tau_2} & \frac{\Delta}{\tau_2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 - \frac{\Delta}{\tau_1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad G = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \frac{\Delta}{\tau_1} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \frac{\Delta}{\tau_1} \\ 0 & 0 \end{pmatrix}$$

A crucial part of the dynamic equations of the arm model is the multiplicative noise structure [Harris and Wolpert, 1998]. Following [Todorov, 2005], control-dependent noise is generated by multiplying the control signal \vec{u}_t with a stochastic matrix and a scaling parameter Σ_u

$$G \Sigma_u \begin{pmatrix} \sigma_t^{(1)} & \sigma_t^{(2)} \\ -\sigma_t^{(2)} & \sigma_t^{(1)} \end{pmatrix} \vec{u}_t$$

Accordingly, the matrices C_i ($i = 1, 2$) are set to

$$C_1 = \begin{pmatrix} \Sigma_u & 0 \\ 0 & \Sigma_u \end{pmatrix} \quad C_2 = \begin{pmatrix} 0 & \Sigma_u \\ -\Sigma_u & 0 \end{pmatrix}$$

Feedback is provided by delayed and noisy measurement of position and velocity of the cursor, and proprioception. The system formulation of equation (1) already implied a feedback delay of one time step, since the sensory feedback y_t is received after generation of the control signal u_t .

Including an additional delay of d time steps can be achieved easily by further augmenting the state space as described in the literature [Todorov and Jordan, 2002]. For the present simulations a feedback delay of $150ms$ was assumed. This yields the feedback equation

$$\vec{y}_t = \left[p_{t-d}^x \quad v_{t-d}^x \quad f_{t-d}^x \quad p_{t-d}^y \quad v_{t-d}^y \quad f_{t-d}^y \right]^T + \chi_t$$

When introducing a parameter uncertainty as in (1), long feedback delays can lead to substantial problems in the process of joint estimation, such as instability and oscillations in the parameter estimate. In fact, the joint estimation of states and parameters can only be accomplished if the parameters change on a time-scale well below the delay time. To circumvent these problems we simply iterated the Kalman equations (4)-(5) at every time step t from $t' = 0$ to $t' = t$ by setting $\vec{a}_{t'} = \vec{a}_t$ and $P_{t'}^a = P_t^a$ for $t' = 0$. This solution is still causal, but makes explicit use of the knowledge that the unknown parameters are constant throughout the control task.

2 Model Fit

For the investigated visuomotor learning task, the cost function J is given by

$$J = \frac{1}{2} \mathbb{E} \left[\sum_{t=0}^{\infty} \left\{ \vec{x}_t^T Q \vec{x}_t + \vec{u}_t^T R \vec{u}_t \right\} \right] \quad (18)$$

with

$$Q = \begin{pmatrix} w_p^2 & 0 & 0 & 0 & -w_p^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & w_v^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -w_p^2 & 0 & 0 & 0 & w_p^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & w_p^2 & 0 & 0 & 0 & -w_p^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & w_v^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -w_p^2 & 0 & 0 & 0 & w_p^2 \end{pmatrix} \quad R = \begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix}$$

Following equation (13), the certainty-equivalent controller then takes the form

$$\vec{u}_t^D = -L_t[\hat{\phi}_t] \vec{x}_t \quad (19)$$

where L_t is computed according to equation (14) and the estimates \vec{x}_t and $\hat{\phi}_t$ are procured by the Unscented Kalman Filter that operates in the augmented state space

$$\vec{\mathbf{x}}_t = \begin{bmatrix} \vec{x}_t \\ \hat{\phi}_t \end{bmatrix} \quad (20)$$

An example of how the parameter estimate evolves within a trial can be seen in **Fig. S3**.

As discussed in the previous section, optimal adaptive controllers generally have to be designed in a problem-specific fashion. To this end, issues of cautious and probing control have to be tackled. In the present case, the probing control problem can safely be neglected, since center-out movements automatically entail better system identification of the rotation parameter $\hat{\phi}$. However, it is intuitively clear that cautiousness (slowing down) is expedient in the presence of very slow parameter identification and high feedback gains. In line with previous work in the engineering sciences [Chakravarty and Moore, 1986; Papadoulis et al., 1987; Papadoulis and Svoronos, 1989], cautiousness is introduced here heuristically by means of an innovation-based “cautious factor”. The basic idea is to tune down feedback gains if the parameter innovation is high, i.e. if the current parameter estimate yields poor predictions. The parameter innovation can be obtained by calculating the Robbins-Munro innovation update [Ljung and Söderström, 1983]

$$I_{t+1}^{\hat{\phi}} = (1 - \alpha) I_t^{\hat{\phi}} + \alpha K_t^{\hat{\phi}} \left[\vec{y}_t - \vec{y}_t^- \right] \left[\vec{y}_t - \vec{y}_t^- \right]^T (K_t^{\hat{\phi}})^T$$

where $K_t^{\hat{\phi}}$ corresponds to the respective entries of the Kalman gain matrix K_t from the Unscented Kalman Filter working on the augmented state space⁵. An example of the innovation estimator can be seen in **Fig. S3**. Importantly, the parameter innovation $I_t^{\hat{\phi}}$ can be used to adjust the feedback gain L_t . In the present case, the feedback gain is effectively determined by the two cost parameters w_p and w_v . They specify the controller’s drive to regulate the position towards the target position, while trying to regulate the velocity to zero. Since the cost function is invariant with regard to a scaling factor (i.e. r can be set arbitrarily), cautiousness can be introduced most generally by means of two effective cost parameters

$$\begin{aligned} \tilde{w}_t^p &= \frac{w_p}{1 + \lambda_p I_t^{\hat{\phi}}} \\ \tilde{w}_t^v &= w_v \left(1 + \lambda_v I_t^{\hat{\phi}} \right) \end{aligned}$$

with constants λ_p and λ_v . While the original cost function is still determined by w_p and w_v , the effective cost parameters \tilde{w}_t^p and \tilde{w}_t^v (i.e. the effective cost matrix \tilde{Q}_t) can be used to calculate the (approximatively) optimal adaptive feedback gain. The optimal adaptive controller then takes the form

$$\vec{u}_t^{opt} = -\tilde{L}_t[\hat{\phi}_t] \vec{x}_t \quad (21)$$

with $\tilde{L}_t[\hat{\phi}_t]$ from equation (14) calculated on the basis of \tilde{Q}_t . In the absence of parameter uncertainty ($I_t^{\hat{\phi}} = 0$) this yields a standard LQG control scheme. For infinitely high innovations one gets $\tilde{w}_t^p \rightarrow 0$ and $\tilde{w}_t^v \rightarrow \infty$, i.e. the controller halts.

Finally, the model was tested on the experimental movement data. To this end, four effective control parameters and three noise parameters of the model were adjusted to fit the mean trajectory and variance of 90°-transformation trials. The parameters of the arm model were taken from the literature [Todorov, 2005]. The obtained parameter set was then used to predict trajectories,

⁵We set $\alpha = 0.1$

speed profiles, angular speed and variance for all intermediary transformation angles and standard movements.

The parameter set to fit and predict the human movement data was as follows:

$$\begin{aligned} \text{Arm Parameters} \quad \tau_1 = \tau_2 &= 40ms \\ m &= 1kg \end{aligned}$$

$$\begin{aligned} \text{Control Parameters} \quad w_p &= 1 \\ w_v &= 0.1 \\ r &= 0.0001 \\ \lambda_p &= 2 \cdot 10^4 \\ \lambda_v &= 1 \cdot 10^4 \end{aligned}$$

$$\begin{aligned} \text{Noise Parameters} \quad \Omega_\xi &= 0 \\ \Omega_\chi &= \left(0.1 \text{diag}([1cm \ 10cm/s \ 100cN \ 1cm \ 10cm/s \ 100cN])\right)^2 \\ \Omega_\nu &= 10^{-7} \\ \Sigma_u &= 0.7 \end{aligned}$$

The average costs appertaining to this parameter set mounted up to $J = 7880 \pm 60$. In contrast, a certainty-equivalent controller ($\lambda_p \equiv \lambda_v \equiv 0$) yields $J^{CE} = 8910 \pm 70$. This clearly shows that for fast movements it is optimal to behave cautiously (cf. **Fig. S5**).

References

- [1] Åström KJ, Wittenmark B (1989) Adaptive Control, Addison-Wesley Publishing Company
- [2] Bar-Shalom Y (1981) Stochastic dynamic programming: caution and probing, IEEE Transactions on automatic control AC-26:1184-1195
- [3] Bar-Shalom Y, Tse E (1974) Dual effect, certainty equivalence, and separation in stochastic control, IEEE Transactions on automatic control AC-19:494-500
- [4] Campi MC (1997) Achieving optimality in adaptive control: the “bet on the best” approach, Proc. 36th Conf. on Decision and Control, pp. 4671-4676
- [5] Campi MC, Kumar PR (1996) Optimal adaptive control of an LQG system, Proc. 35th Conf. on Decision and Control, pp. 349-353
- [6] Chakravarty A, Moore JB (1986) Aircraft flutter suppression via adaptive LQG control, Proc. American Control Conf., pp. 488-493
- [7] Harris CM, Wolpert DM (1998) Signal-dependent noise determines motor planning, Nature 394:780-784
- [8] Haykin S (2001) Kalman filtering and neural networks, John Wiley and Sons, Inc., New York
- [9] Jazwinsky A (1970) Stochastic processes and filtering theory, New York: Academic Press
- [10] Julier SJ, Uhlmann JK, Durrant-Whyte H (1995) A new approach for filtering nonlinear systems, Proc. Am. Control Conference, pp. 1628-1632
- [11] Kalman RE (1960) A new approach to linear filtering and prediction problems, Transactions of ASME, Ser. D, Journal of basic engineering 82:34-45
- [12] Kleinman D (1969) Optimal stationary control of linear systems with control-dependent noise, IEEE Transactions on automatic control AC-14(6):673-677
- [13] Kumar PR (1983) Optimal adaptive control of linear-quadratic-Gaussian systems, SIAM J. Control and Optimization 21(2):163-178
- [14] Kumar PR (1990) Convergence of adaptive control schemes using least-squares parameter estimates, IEEE Trans. on Automatic Control, AC-35(5):416-424
- [15] Moore JB, Zhou XY, Lim AEB (1999) Discrete time LQG controls with control dependent noise, Systems & Control Letters 36:199-206
- [16] Papadoulis AV, Svoronos SA (1989) A MIMO cautious self-tuning controller, AIChE 35:1465-1472
- [17] Papadoulis AV, Tsiligiannis CA, Svoronos SA (1987) A cautious self-tuning controller for chemical processes, AIChE 33:401-409

- [18] Sastry S, Bodson M (1989) Adaptive control. Stability, convergence, and robustness, Prentice-Hall Information and System Sciences Series
- [19] Stengel RF (1994) Optimal control and estimation, Dover Publications
- [20] Sutton RS, Barto AG (1998) Reinforcement Learning, MIT Press, Cambridge, Massachusetts
- [21] Todorov E (2005) Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system, *Neural Comp.* 17:1084-1108
- [22] Todorov E, Jordan MI (2002) Optimal feedback control as a theory of motor coordination, *Nat Neurosci.* 5:1226-1235
- [23] van Schuppen JH (1994) Tuning of Gaussian stochastic control systems, *IEEE Trans. on Automatic Control*, AC-39:2178-2190
- [24] Winter DA (1990) Biomechanics and motor control of human movement, John Wiley and Sons, Inc., New York

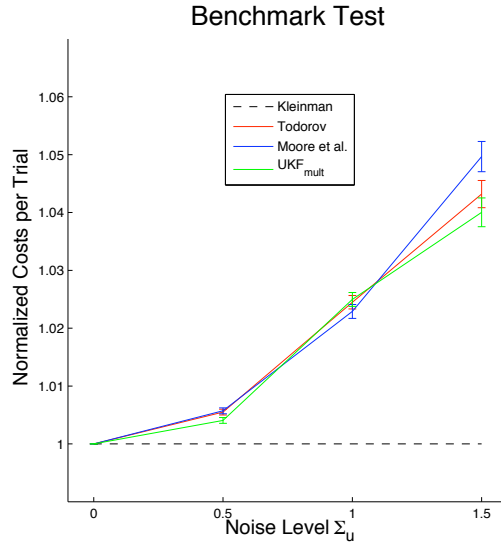


Figure S1. Benchmark Test. The performance of the proposed algorithm (“ UKF_{mult} ”) was measured for different magnitudes of multiplicative noise in a standard center-out reaching task to allow for comparison with existing approximation schemes by [Todorov, 2005] and [Moore et al., 1999]. In the absence of observation noise, Kleinman [Kleinman, 1969] calculated the optimal solution to the posed control problem and, thereby, provided a lower bound. All other algorithms are run in the presence of observation noise and their average costs per trial are normalized by this lower bound. All three algorithms achieve roughly the same performance.

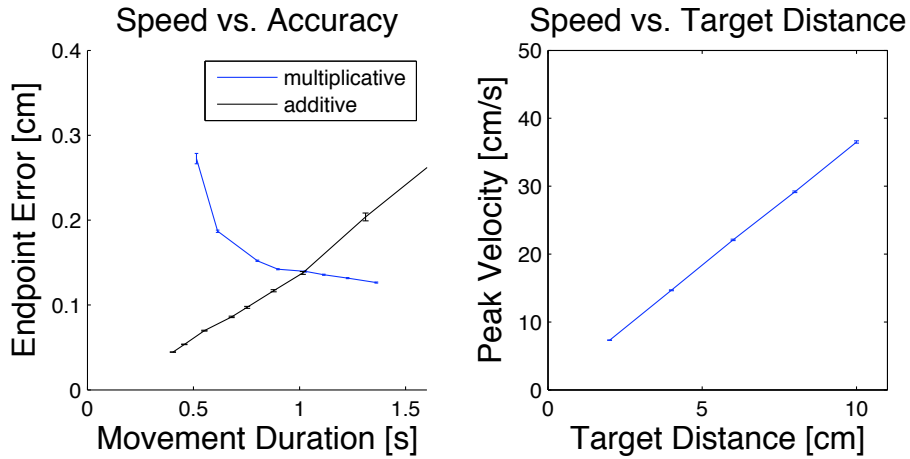


Figure S2. Speed-Accuracy Trade-off. Due to control-dependent multiplicative noise, fast movements entail higher inaccuracy as measured by positional standard deviation once the target is reached. This relationship cannot be explained in an additive noise scenario. Model predictions are in line with the model of Todorov [Todorov, 2005]. (b) Speed vs. Target Distance. The model predicts a linear relationship between target distance and peak velocity as found experimentally. (cf. [Krakauer et al., 2000, Fig. 2D]).

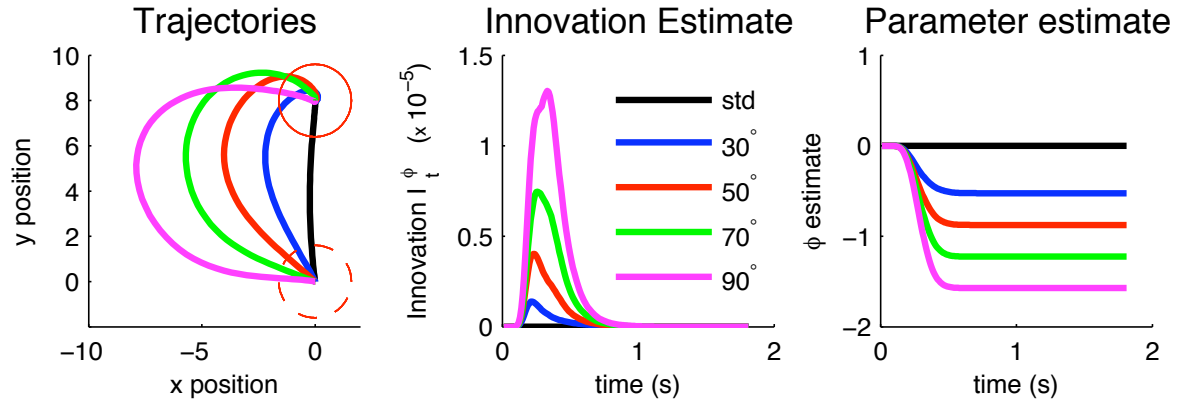


Figure S3. Adaptation of model parameters. The left panel shows trajectories when the controller adapts to different unexpected visuomotor transformations: 0° black, 30° blue, 50° red, 70° green and 90° magenta. The middle panel shows how the innovation estimate evolves within a trial. Due to feedback delay, initially there is no mismatch detected. After the delay time the innovation estimator detects parameter mismatch. Once the correct parameter estimate can be achieved innovations return to zero again. The right panel shows evolution of the parameter estimate within a trial. The different rotation angles are estimated corresponding to different experienced visuomotor rotations.

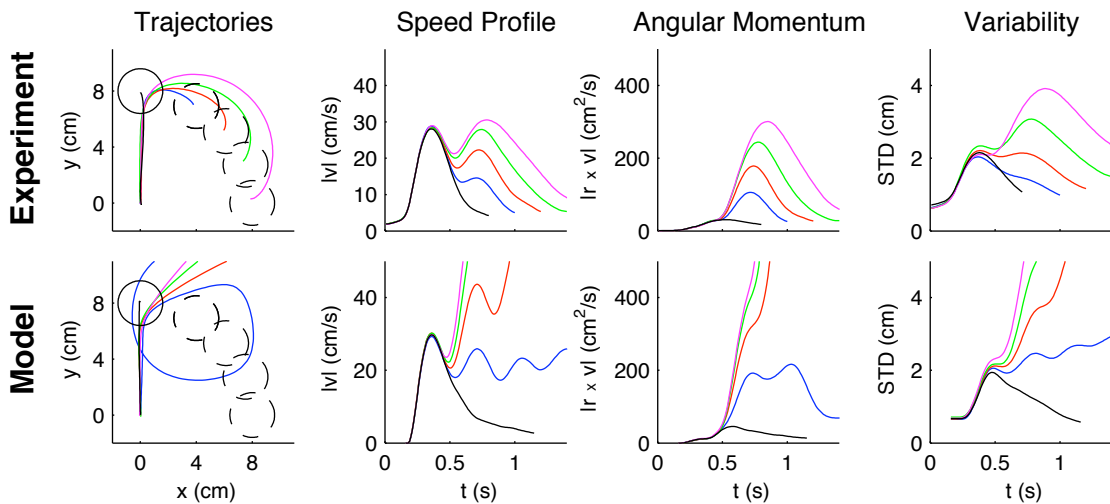


Figure S4. Non-adaptive optimal control model. When the model is not allowed to track the rotation parameter, the controller becomes quickly unstable. The trajectories diverge.

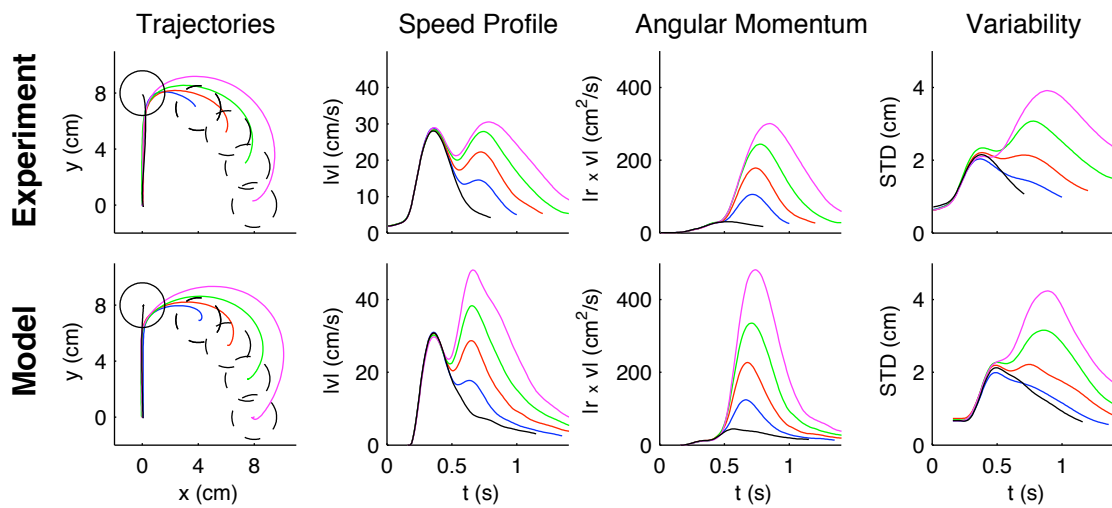


Figure S5. Adaptive controller without “cautiousness”. When the cautiousness parameters are set to zero, the controller acts much faster in the second part of the movement, not only leading to higher speeds, but importantly also to higher costs (compare Section 2).