Title Page

Title: Optimal control of reaching includes kinematic constraints

Running Head: Optimal control of reaching includes kinematic constraints

Authors: Michael Mistry\textsuperscript{1,2}, Evangelos Theodorou\textsuperscript{2}, Stefan Schaal\textsuperscript{1,2}, Mitsuo Kawato\textsuperscript{1}

Author Contributions:
M. Mistry: Performed experiments and analyzed data
M. Mistry and E. Theodorou: Drafted manuscript
All Authors: Conceived and designed research, interpreted results of experiments, edited and revised manuscript

Author Affiliations:
\textsuperscript{1}ATR Computational Neuroscience Labs, Kyoto, Japan
\textsuperscript{2}University of Southern California, Dept. of Computer Science, Los Angeles, CA

Contact Information:
Michael Mistry
m.n.mistry@bham.ac.uk
School of Computer Science, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK
Tel: +44 121 414 8477
Fax: +44 121 414 4281
Abstract

We investigate adaptation under a reaching task with an acceleration-based force field perturbation designed to alter the nominal straight hand trajectory in a potentially benign manner: pushing the hand off course in one direction before subsequently restoring towards the target. In this particular task, an explicit strategy to reduce motor effort requires a distinct deviation from the nominal rectilinear hand trajectory. Rather, our results display a clear directional preference during learning, as subjects adapted perturbed curved trajectories towards their initial baselines. We model this behavior using the framework of stochastic optimal control theory and an objective function that trades-off the discordant requirements of 1) target accuracy, 2) motor effort, and 3) kinematic invariance. Our work addresses the underlying objective of a reaching movement, and we suggest that robustness, particularly against internal model uncertainly, is as essential to the reaching task as terminal accuracy and energy efficiency.

Keywords

motor control, optimal control, reaching, force fields, motor adaptation
Introduction

Optimization has long been suggested as a guiding principle behind sensorimotor control, planning, and execution. Associated with optimization is always a cost function: the minimization of which defines the “best” possible movement or control strategy. Various cost functions have been proposed for human reaching: the minimization of jerk (Flash and Hogan, 1985), torque change (Uno, Kawato, and Suzuki, 1989), variance (Harris and Wolpert, 1998), interaction torques (Goble et al., 2007), or combinations of such (Ohta et al., 2004). Recent models using the framework of stochastic optimal control (SOC) have demonstrated how human reaching behavior can result from a control strategy that minimizes the sum of motor commands during a movement and positional error at the end of the movement (Todorov and Jordan, 2002). Since the cost functions typically employed in these models do not include kinematic constraints, such as a desired trajectory for the hand to the target, they suggest the possibly that the central nervous system (CNS) need not be concerned about hand trajectory to plan and execute a successful reaching movement.

However, an important question to consider is what happens during the adaptation process? When learning to reach in a novel dynamic environment (such as a rotating room (Lackner and DiZio, 1994) or perturbing robot manipulandum (Shadmehr and Mussa-Ivaldi, 1994)), what are the factors that drive the CNS’s adaptation strategy? If we are to believe that optimization guides this process, then the reduction of cost, on a trial-by-trial basis, should be the preeminent motivation. Investigating the concept of “adaptation as reoptimization,” (Izawa et al., 2008) proposed a framework whereby learning occurs via progressively improved estimations of internal models, and subsequent reoptimizations based on those improved models. Additionally, because this procedure may result in a post-adaptation trajectory that differs significantly from baseline, the authors of this work claim that the CNS cannot be optimizing to restore an invariant trajectory. Rather, the CNS adapts only for the sake of improving endpoint accuracy and reducing motor effort.

While accuracy and effort are certainly relevant criteria for the sensorimotor control system, there has been evidence that trajectory does play a significant role in adaptation. For example in (Wolpert, Ghahramani, and Jordan, 1995) subjects received artificially curved visual feedback of their hand position during a reaching task, that did not affect accuracy at the target. Subjects adapted by curving their hand trajectory in the opposite direction in order to maintain a visually straight path, effectively increasing
motor effort. Similarly, (Kistemaker, Wong, and Gribble, 2010) created a novel force field where the minimal energy trajectory differed substantially from the nominally straight hand movement. Even after practicing on the minimum energy path, subjects still chose to restore to the sub-optimal straight trajectory.

In this work, we test how subjects react to a force field perturbation where the maintenance of a rectilinear hand trajectory actively conflicts with the desire to reduce motor effort. We ask how the CNS trades-off the discordant requirements of 1) target accuracy, 2) motor effort, and 3) kinematic invariance. We suggest how the simple inclusion of kinematic constraints in the stochastic optimal control framework unifies this powerful theory with previous kinematically motivated models. We also suggest that such constraints are a means to robustify the neural controller against environmental uncertainty, to aid in task achievement while bootstrapping the process of exploration and learning.

Materials and Methods

A total of sixteen healthy, right-handed subjects (ages 20-32, fourteen male, two female), participated in this study. The institutional ethics committee of ATR Institute International approved this experiment, and subjects gave informed consent.

Inspired by the visuomotor adaptation experiment of (Wolpert, Ghahramani, and Jordan, 1995), we attempt to create a dynamic environment that affects motor cost, but not accuracy: while reaching to a goal, if a force perturbation first pushes the hand off-course in one direction and then subsequently back in the opposite direction, the goal may still be achieved with minimal correction. Fighting this perturbation in order to maintain a rectilinear hand trajectory may unnecessarily increase motor effort. Although there are potentially many such force-fields that can accomplish this directive (e.g. (Kistemaker, Wong, and Gribble, 2010)), we also wished to model such a disturbance within existing linear optimal control modeling techniques. Therefore we chose a perturbation that is a linear function of hand kinematics:

\[ F_x = 2.0 \ddot{y} - 5.0 \dot{x} \]

Thus, if the subject plans to execute a straight, minimum jerk-like movement, a disturbing force will first push the hand off course to the right during acceleration, and then subsequently left, back towards the
target during deceleration (Figure 2). The viscous term is required to dampen inertia sufficiently such that the deceleration disturbance actually reverses hand direction, and additionally contributes to the overall stability of the device. The gains of the force field were tuned in pilot studies (results not shown) with the general goal of returning the hand to the target after a substantial initial off-course deviation (at least after the first “catch-trial” of force field onset, before adaptation occurs).

Additionally, because (Wolpert, Ghahramani, and Jordan, 1995) already addressed the issue of visually perceived error, we focus solely on the case of proprioceptive feedback. Thus we give no visual feedback to our subjects, except for the terminal position (hand position at the end of a trial). Also, because we wanted to focus our study to the adaptation of the acceleration term we left the viscous damping term on for all “null” field trials.

Apparatus

We use the ATR parallel-link, 2 degree-of-freedom, direct drive, air-magnet floating manipulandum (PFM), shown in Figure 1. Subjects were seated in a chair and strapped in with shoulder harness to keep their back against the chair. A wrist brace was attached to their right hand to restrict wrist motion and connect the subject to the handle of the device. The combination of wrist brace and shoulder straps only allow for planar shoulder and elbow motion. The subjects cannot see their arm, as the tabletop covers the entire workspace. Start and target locations are projected onto the table, with a ceiling mounted projector, as well as the positional feedback at the end of the movement. Kinematic data is recorded at 500Hz rate. The force applied by subjects at the handle of the PFM is measured with a force sensor (Nitta Corp. no. 328) with resolution of 0.06 N. The manipulandum and setup were described in detail previously (Gomi and Kawato, 1996).

Experimental Procedure

Subjects are asked to execute 25cm point-to-point reaching movements, within a time window of 800+/−150ms, to a specified target in a single direction, perpendicular to the frontal plane, away from the body. The target is indicated by a 2.5cm diameter circle. Before the start of each trial, the subjects move their hands to the starting location, indicated by a 2cm diameter circle. During this phase, visual feedback of the hand position is given (indicated by a dot), but only within 4 cm of the starting location, and only so the subject can properly position their hand to start. Once inside the starting circle, visual feedback is
removed, and 3 consecutive beeps sound to indicate the start of the trial. Subjects are asked to initiate movement after conclusion of the 3 beeps, and stop at the target within the specified time window. No other instructions regarding how to execute the movement or which path to take to the target were given. No visual feedback is given during the trial, except a single dot indicating the final hand position at the end of the trial. The subject is also told after each trial whether the trial was a success or failure, where success is defined by the hand arriving and stopping within the target circle during the specified time window. If the trial was a failure, the subject is told if the hand arrived too late, too early, or was off target.

**Experimental Conditions**

Each subject was tested under one of two conditions. Under the first condition (Experiment A), after initial training in the null field, the subjects experienced the full strength of the force field within a series of random catch trials. For this group of subjects, we were able to estimate the Before Effects: reaching behavior in the force field before adaptation onset. As the Before Effects were strongly curved trajectories, significantly different from baseline, we were concerned that subjects would be aware of the abrupt change to motor behavior, and subsequently make a conscious effort to adapt their motor strategy. In order to minimize the effect of conscious adaptation, we also tested a separate group of subjects under a condition where force field strength gradually increased over a series of trials (Experiment B). In both conditions, after sufficient training in the force field we were able to measure After Effects (by switching the force field off in random catch trials).

Experiment A had 6 subjects (5 male, 1 female). Each was tested in 5 blocks of trials, with a 5 minute break in between each block. In Block 1 subjects trained on the null field condition until 70 successful trials were completed. In Block 2, subjects executed 30 successful null field trials, followed by 80 trials containing 10 random catch trials (where the field is switched on unknowingly to the subject). Blocks 3 and 4 required 60 successful trials each in the force field. Finally, in Block 5 subjects executed 30 successful force field trials followed by 80 trials containing 10 random catch trials (with the field off during these 10 trials) to test for After Effects. Over the course of the session, each subject trains for a total of 100 successful null field trials, and 150 successful force field trials. For Experiment B, we tested a total of 10 subjects (9 male, 1 female) with 4 blocks (again with 5 minute breaks in between each block). Block 1 was identical as in Experiment A. In Block 2, subjects executed 30 successful null field trials,
followed by 18 trials with the force field gain linearly rising from trial to trial until reaching full strength, and then followed by an additional 40 successful trials in the full strength force field. Block 3 required 70 successful trials in the full strength force field. Finally, in Block 4 subjects executed 40 successful trials in the force field followed by 80 trials containing 10 random catch trials with the field off. Similarly to Experiment A, each subject trains for a total of 100 successful null field trials, and 150 successful force field trials.

**Data Analysis**

In order to quantify the curvature of a trajectory, we compute the average perpendicular distance among all points in the trajectory from a straight line that connects the start of the trial to the target. We assign this metric a positive value if the greater portion of the trajectory is on the right side of the line, otherwise it receives a negative value. For the purpose of this data analysis, the start of a trial is considered to be the last time the velocity is 0.05m/s before reaching peak velocity. The end of the trial is the first time the velocity reaches 0.05m/s after peak velocity.

**Modeling and Simulation**

Stochastic optimal control (SOC) has been a powerful tool for modeling the estimation and control of biological sensorimotor systems, particularly because such systems are inherently noisy and partially observable (Loeb, Levine, and He, 1990). If one assumes a linear model of dynamics with additive Gaussian noise, and a globally quadratic cost function, a cost minimizing optimal controller can be derived analytically via the Linear-Quadratic-Gaussian (LQG) framework. Additionally, (Todorov, 2005) extends the LQG framework to handle the case of signal-dependent noise, a property believed to exist in biological muscles and affect sensorimotor planning (Harris and Wolpert, 1998).

**Stochastic Optimal Control**

The SOC model, in its general form, assumes the following linear dynamics equation:

\[ x_{t+1} = Ax_t + Bu_t + \zeta_t + \sum_{i=1}^{c} \epsilon_i C_i u_t \]  

(2)
where $x_t$ is the state vector, $u_t$ is the control input vector, and $\xi_t$ and $\epsilon_t^i$ are the additive and multiplicative zero-mean Gaussian noise variables, respectively. $A$ is the state transition matrix, $B$ the control input matrix, and $C_i$ are scaling matrices for control dependent noise. The model also assumes sensory feedback is partial observable, and given by the equation:

$$y_t = Hx_t + \omega_t$$

where $y_t$ are the observations, $H$ is the observation matrix, and $\omega_t$ is zero-mean Gaussian noise. Additionally, the following quadratic cost is accrued at each time step:

$$0 \leq x_t^TQ_t x_t + u_t^TR u_t$$

where the matrix $R$ is a positive definite, and the matrix $Q_t$ is positive semi-definite. Under these assumptions, (Todorov, 2005) derives the optimal controller and estimator in the following form:

$$u_t = -L_t \hat{x}_t$$

$$\hat{x}_{t+1} = (A - BL_t) \hat{x}_t + K_t (y_t - H\hat{x}_t) + \eta_t$$

where $\hat{x}_t$ is the state estimate at time $t$, and $\eta_t$ is zero mean Gaussian noise. The optimal control gain, $L_t$, and optimal estimator gain (or Kalman gain), $K_t$, are computed via an iterative procedure derived from dynamic programming principles. Please see (Todorov, 2005) for full details.

**A model of sensorimotor control of human reaching**

The assumptions of linear dynamics and a globally quadratic cost may seem to be too restrictive, particularly for the highly nonlinear problem of sensorimotor control of human reaching. However, as demonstrated in (Todorov, 2005), SOC still has a predictive power, particularly for systems operating under uncertainty. Following the model of the motor system derived in (Todorov, 2005), and also employed in (Izawa et al., 2008), we represent the human arm with a 10-dimensional state vector as follows:

$$x_t = \left( p_t^T \ v_t^T \ f_t^T \ g_t^T \ p^T \right)^T$$
where $\mathbf{p}_t$ and $\mathbf{v}_t$ are the 2-dimensional position and velocity of a point mass, respectively. The vector $\mathbf{p}^*$ represents the $x$ and $y$ target location. Muscles are modeled as a second order linear system, where their output force, $\mathbf{f}_t$, is a function of muscle activity, $\mathbf{g}_t$, and muscle activity is driven by control input with signal-dependent noise. The complete linear dynamics equations of the arm model are defined as follows:

$$
\mathbf{p}_{t+1} = \mathbf{p}_t + \mathbf{v}_t \Delta t \quad (8)
$$

$$
\mathbf{v}_{t+1} = \mathbf{v}_t + \left( \frac{\mathbf{f}_t + D\mathbf{x}_t}{m} \right) \Delta t \quad (9)
$$

$$
\mathbf{f}_{t+1} = \mathbf{f}_t \left(1 - \frac{\Delta t}{\tau_2}\right) + \mathbf{g}_t \frac{\Delta t}{\tau_2} \quad (10)
$$

$$
\mathbf{g}_{t+1} = \mathbf{g}_t \left(1 - \frac{\Delta t}{\tau_1}\right) + \mathbf{u}_t (1 + \sigma_c \varepsilon_t) \frac{\Delta t}{\tau_2} \quad (11)
$$

where $\Delta_t$ is the time step and $D\mathbf{x}_t$ (used in (9)) is some linear, state dependent, external force, such as a force field. For example, in the case of our acceleration based force field (1), we write:

$$
D\mathbf{x}_t = \beta \begin{bmatrix}
0 & 2.0/m \\
0 & 0
\end{bmatrix} \mathbf{f}_t - \begin{bmatrix}
5.0 & 0 \\
0 & 0
\end{bmatrix} \mathbf{v}_t. \quad (12)
$$

The parameter $\beta$ is used in the modeling of Experiment B to gradually scale field strength from 0 to 1. Otherwise, $\beta = 1$ when the field is on, and $\beta = 0$ in the null field case.

The total quadratic cost, accumulated for the entire movement, is defined as follows:

$$
w_p (\mathbf{p}_{t_f} - \mathbf{p}^*)^T (\mathbf{p}_{t_f} - \mathbf{p}^*) + w_v \mathbf{v}_{t_f}^T \mathbf{v}_{t_f} + w_f \mathbf{f}_{t_f}^T \mathbf{f}_{t_f} + \sum_{t=0}^{t_f} \mathbf{u}_t^T \mathbf{R} \mathbf{u}_t, \quad (13)
$$

or in a more compact notation:

$$
\mathbf{x}_{t_f}^T Q_{t_f} \mathbf{x}_{t_f} + \sum_{t=0}^{t_f} \mathbf{u}_t^T \mathbf{R} \mathbf{u}_t, \quad (14)
$$

where the parameters $w_p$, $w_v$, $w_f$, and $R$ weight the importance of target accuracy, terminal velocity, terminal force, and control input respectively. This cost function effectively defines the objective of the reaching movement: be at the target by a specified time ($t_f$), and stop there (with zero velocity and zero force), all while using the minimal amount of control effort as possible. It is important to note that the cost function does not specify any direction or path to the target. Assuming weight parameters are
appropriately tuned, various human-like motions (e.g. minimum jerk trajectories) are shown to emerge from the minimization of this trade off between target accuracy with motor effort (Todorov, 2005). For reasons that will become apparent, we henceforth refer to this cost function as \textit{non-directional}.

\textbf{Reoptimization as a model of learning}

The work of (Izawa et al., 2008) extends the SOC framework to incorporate learning and uncertainty of force field dynamics. Here, we will borrow their technique of modeling learning behavior. When learning a new force field, the subject does not have full knowledge of the external force matrix $D$. Rather, the subject can only estimate $D$ as:

$$\hat{D} = \alpha D,$$  \hspace{1cm} (15)

with $0 \leq \alpha \leq 1$. Thus, learning behavior can be approximated by gradually increasing $\alpha$, effectively improving the subject’s estimation of the force field. When implementing this framework, the optimal control and estimator gains are computed using the estimate, $\hat{D}$, while the resulting simulation uses the actual $D$. In this work, we also use reoptimization to model behavior during catch trials. Before Effects are modeled simply as $\beta = 1.0$ and $\alpha = 0.0$. To model After Effects we use $\beta = 1.0$ and $\alpha = 1.0$ in the model estimate, however when running the simulation we set $\beta = 0.0$.

\textbf{Modeling directional constraint}

In this work, we also extend the SOC cost function (14) to include a kinematic constraint\footnote{Here we treat kinematic invariance as a \textit{soft} constraint, i.e. as an optimization objective, as opposed to a \textit{hard} constraint which must necessarily be satisfied by the optimization routine.}, or preferred direction of motion. Calling $d = [d_x \hspace{0.5cm} d_y]^T$ a unit vector representing desired direction, we define the matrix $Q_d$ as:

$$Q_d = \begin{bmatrix} d_y^2 & -d_x d_y \\ -d_x d_y & d_x^2 \end{bmatrix}$$  \hspace{1cm} (16)

and add new terms to the original cost function, with the intent to penalize any position or velocity perpendicular to the desired direction:

$$x_{t_j}^T Q_d x_{t_j} + \sum_{t=0}^{t_i} \left( u_t^T R u_t + e^{-t \Delta t / \tau} \left( k_p p_t^T Q d p_t + k_v v_t^T Q d v_t \right) \right),$$  \hspace{1cm} (17)
where $k_p$ and $k_v$ weight the importance of error in position and velocity, respectively. An exponential decay term is also included to emphasize that the kinematic constraint need not exist for the entire motion. Note that the new cost function remains quadratic with respect to the state vector, and thus can be used in LQG methods. For the purpose of our experiments and simulations, we set $d = [0 \ 1]^T$ (the direction towards the target at the start of the motion), and the cost function simplifies to:

$$x_t^T Q_t x_t + \sum_{t=0}^{t_f} \left( u_t^T R u_t + e^{-t \Delta_t / \tau} \left( k_p p_x^2 + k_v v_x^2 \right) \right).$$

(18)

Simulation examples with a viscous curl field

For verification purposes, we test the SOC with reoptimization framework, under a traditional rightward viscous curl field. We use the same target distance and time constraints as our intended experiment, and tune the parameters of the non-directional cost function to achieve qualitatively the same performance as in (Izawa et al., 2008). Figure 3 (left) shows the trajectory predictions made by the SOC framework with increasingly improved estimates of the force field. This demonstration repeats and confirms the results reported in (Izawa et al., 2008), and additionally shows our predictions of before and after effects. Note that, as emphasized in (Izawa et al., 2008), the fully adapted optimal controller does not move straight to the target, but rather moves with a significant leftward curvature. Less total effort can be expended by first moving against the field, while velocity and force are low, and subsequently allowing the field to push the hand back towards the target. We add directional constraint to the cost function, with $\tau = 0.07$ and tune only the directional constraint terms to achieve a nearly straight trajectory in the fully adapted case. All other parameters are identical to the previous case. Not surprisingly, using a directional constraint can result in an optimal trajectory close to baseline (Figure 3 (right)).

Results

Figure 4 (rightmost) shows the trajectory plots averaged across all 6 subjects of Experiment A (A.DATA), as well as the average across the 10 subjects of Experiment B (B.DATA). The subjects’ nearly-straight baseline trajectory is the result after at least 100 successful trails in the null field. The significant rightward curvature seen in the Before Effect (A.DATA) illustrates the preadaptation effect of the acceleration force field perturbation on the reaching movement: pushing the hand to the right before restoring towards the
target. Final Force trajectories show subsequent adaptation after at least 150 successful training trials in the force field. Subjects either restored (in the case of Experiment A) or maintained (in the case of Experiment B) their hand trajectories to approximately baseline. Strong leftward curved After Effects (in both conditions) indicate the formation of a predictive internal model of the force field dynamics. The thin lines in A.DATA and B.DATA show some incremental stages of learning (the statistics of which are reported in Tables 1 and 2). The learning curve plots in Figure 5 additionally show how average curvature progresses during force field training. Figure 6 DATA are the force measurements in the direction perpendicular to the target (same direction as the force field perturbation), recorded by the force sensor at the handle of the manipulandum. Both the averaged force trajectories (averaged across all 6 subjects of Experiment A) as well as the individual trajectories from a single subject, have substantial leftward (negative) force, early in the movement, providing resistance against the force field perturbation. Evidence of resistive forces in preadaptation trajectories (Before Effects) indicate that the opposing force may be reactionary, correcting for error observed within a single trial, and not necessarily preplanned before execution.

**Modeling using stochastic optimal control**

Using the linear dynamics model described previously and the non-directional cost function defined by (14), we use the stochastic optimal control framework to compute the optimal control policy under our force field. Cost function parameters were tuned to achieve bell-shaped velocity profiles and goal achievement in the baseline and fully-adapted cases. Not surprisingly, the SOC model predicts that the optimal behavior in the force field is to move with a significant rightward curvature, allowing the disturbance to push the hand off course before restoring towards the target (A.SIM.nd: Final Force). The preadaptation trajectory is curved to the right, but misses the target towards the left (A.SIM.nd: Before Effect). The reoptimization model predicts increasingly outwards curvature as better estimates of the force field are obtained (A.SIM.nd: thin lines between Before Effect and Final Force). Increasingly outwards curvature is also predicted if field strength is gradually increased from zero (B.SIM.nd: thin lines between Baseline and Final Force). Results of these simulations imply that the stochastic optimal control and reoptimization framework fail to predict the correct adaptation behavior of our subjects.

We add the directional constraint term to the cost function as in (18), and repeat the simulations. We set the time constant of the decay term to $\tau = 0.07$ to achieve 90% decay at approximately the
time of peak acceleration (see Discussion). We additionally tune only the directional constraint terms of the cost function to change the adaptation behavior, and better represent results from our subjects. All other parameters remained the same as in the non-directional case. Using a directional cost function, reoptimization now predicts an inward adaptation trend (A.SIM.d: thin lines between Before Effect and Final Force), until the final trajectory is close to Baseline (A.SIM.d: Final Force). Gradually increasing field strength from zero, results in little change to curvature (B.SIM.d). Output forces from the directional cost model (Figure 6 SIM) resist the force field, matching the behavior observed in subjects. On the contrary, the non-directional model applies force in the same direction as the force field at the initial stage of the movement. Overall, the SOC model with a directional cost function provides an improved resemblance to our behavioral results.

Discussion

We investigated adaptation within a reaching task where perturbations altered nominal hand trajectory in a potentially benign manner: pushing the hand off course in one direction before subsequently restoring towards the target. The task was designed to elicit the underlying adaptation strategy employed by the CNS, since in this particular task, a strategy to reduce motor effort requires a distinct deviation from a rectilinear hand trajectory. Rather, our results display a clear directional preference during adaptation. Hypotheses proposing that kinematic error does not play a role in sensorimotor adaptation, e.g. as suggested by (Izawa et al., 2008), cannot hold.

Although our results concur with a long standing viewpoint that kinematics influences motor behavior, it remains clear that strict invariants such as minimum jerk or rectilinearity do not dominate motor control, e.g. as proposed in older works such as (Flash and Hogan, 1985) and (Shadmehr and Mussa-Ivaldi, 1994). Extended trials in viscous force fields resulted in an adaptation to energy optimal curved trajectories (Izawa et al., 2008), especially without visual feedback (Arce et al., 2009). We also speculate that with sufficient training in our acceleration force field, subjects may also learn to conduct curved trajectories. Indeed, the reduction of motor effort is a powerful motivation for the CNS. In this work, however, we address the short term adaptation strategy, particularly during periods of significant environmental uncertainty. We argue that during adaptation, a robust strategy can be implemented via a control mechanism that trades off the desire for a rectilinear trajectory with the aspirations of terminal
accuracy and low motor effort. In our model, we propose an optimal control cost function that represents this trade-off. Adding a kinematic constraint term to the stochastic optimal control (SOC) cost function introduces the additional requirement of rectilinearity while still able to exploit the predictive power of SOC for modeling biological systems with signal dependent noise and partial observability. Thus, we are able to unify the conflicting notions of desired trajectory and energy efficiency within an existing computational framework.

There has been a long standing discourse in motor control literature regarding the desired trajectory hypothesis (DTH), i.e. the notion that a movement trajectory is preplanned before execution. DTH has largely been supported by observations that subjects restored baseline trajectories after adaptation (Shadmehr and Mussa-Ivaldi, 1994), (Wolpert, Ghahramani, and Jordan, 1995). DTH is also appealing from a computational perspective, as trajectory tracking is a simple, well studied control methodology, commonly used in engineered systems. Recently, however, significant works have used stochastic optimal control theory to explain the trajectory as well as variability profiles of human reaching by means of only parsimonious task criteria: maximization of target accuracy and minimization of control effort. The main rationale behind these works is dictated by the minimum intervention principle (Todorov and Jordan, 2002): deviations from baseline need only be corrected if they interfere with task performance. If the task is only to minimize energy and maximize end-point accuracy, prescribing a trajectory before execution makes little sense, as particular paths to the target are often inconsequential to goal achievement. Moreover, the process of enforcing such a trajectory may add extraneous motor cost, and in the case of systems with signal dependent noise, introduce additional execution noise. Rather, the sensorimotor system is better left unrestrained by kinematic invariants, and able to reserve decisions of how to best reach the target at execution time, when relevant contextual information is available.

This rational makes sense for the nominal case of learned reaching movements, where underlying dynamics are fully known, and can be exploited to achieve the task accurately and efficiently. However, during a learning scenario, e.g. when exposed to a novel force field, the use of an inaccurate model becomes detrimental and potentially unsafe. In this case, to successfully reach the target, the brain must employ a robust strategy that is less sensitive to modeling inaccuracies. Any unknown force field dynamics can be treated as disturbances to be rejected, at least until an improved model is formed. Humans often apply such a strategy during learning by increasing arm stiffness, thereby reducing the effect of unknown disturbances (Takahashi, Scheidt, and Reinkensmeyer, 2001), then subsequently lowering stiffness levels
as a new model is formed (Franklin et al., 2008). In our model, we see that adding kinematic constraints to the objective function increases the position and velocity gains of the controller, perpendicular to target direction, during the early part of the movement before the constraint has decayed (Figure 7). For our linear system, these gain increases can effectively be viewed as stiffness and damping increases. The amplification of gains is likely to result in more overall energy expenditure, however, but is also likely to reduce the deviations from target direction. We see this effect when we compute the controller using the nominal mass value, but test on varying mass values. A controller with directional constraint will be less sensitive to erroneous internal models (Figure 8). Thus we argue that kinematic invariants, such as a directional preference, do more for the CNS than merely simplifying the control problem or improving its computational tractability. Kinematic constraints serve to robustify the neural controller against the uncertain environment, to aid the arm in reaching the target, while bootstrapping the process of exploration and learning.

In order to be effective, these kinematic constraints need not be required for the full length or duration of the movement, dictating a complete trajectory from start to goal. A vectorial position and velocity constraint, affecting only the initial part of the movement, is consistent with the idea that reaches are vectorially planned (Bullock and Grossberg, 1989), (Bock and Arnold, 1992), (Gordon, Ghilardi, and Ghez, 1994), (Ghilardi, Gordon, and Ghez, 1995), (Krakauer et al., 2000), i.e. planned in a hand centered coordinate system, with a desired initial direction of movement. For example, (Gordon, Ghilardi, and Ghez, 1994) observed that directional variability at the start of a movement was constant over several directions and independent of distance, speed, and inertial properties of the limb, concluding that the directional variability is mostly due to errors in the planning process. The work of (Ghilardi, Gordon, and Ghez, 1995) also showed that hand trajectories had systematic directional biases, in the absence of vision, even when hands where repositioned to a new workspace, and (Ghez, Scheidt, and Heijink, 2007) showed directional biases in slicing task (out and back in one continuous motion) when initial hand positions were covertly altered by a robotic manipulandum.

Our model of directional preference also includes a time dependent decay. Although a constant, non-decaying, directional constraint term would achieve qualitatively the same outcome (results not shown), including the decay stresses that the directional constraint has significant influence at the start of the movement. Reduction of the constraint during motion is consistent with the notion that initial trajectory and endpoint accuracy are controlled sequentially (Scheidt and Ghez, 2007). In that work, subjects who
trained on visuomotor rotations with slicing movements did not transfer learning to a reaching task (and visa-versa). They suggested that separate control mechanisms may exist for initial trajectory and final posture, rather than a complete spatial plan from start to goal. In their non-linear model using inverse dynamics, the two controllers act sequentially with onset of the posture controller occurring at the time of peak acceleration. Motivated by this idea, we set the time constant of the decay term to achieve 90% decay at approximately the time of peak acceleration. Our results demonstrate how the concept of “sequential” control can be encapsulated within optimal control theory, by trading off requirements in a cost function, instead of relying on a switching or blending of separate controllers.

Finally, there are certain physiological limitations that may inhibit the process of learning a novel dynamics environment. Evidence suggests that internal model formation occurs locally, and does not generalize to untrained areas of the workspace for both dynamic perturbations (Gandolfo, Mussa-Ivaldi, and Bizzi, 1996) and visuomotor transformations (Krakauer et al., 2000). Thus, for complete learning, the workspace must be exhaustively explored, which can be time consuming, energetically expensive, and risky. We also wish to learn to make energy efficient movements, however, our sensory system has generally poor feedback of instantaneous motor cost. Feedback of energy consumption can be substantially delayed, perhaps on the time course of hours or days, in terms of muscle fatigue or exhaustion, e.g. pain in response to lactic acid (Allen, Lamb, and Westerblad, 2008). In the short term, however, a perceived deviation from a kinematic invariant can be relied upon as indication that something has gone awry with the current internal model of the dynamical system. Then, a strategy that systematically attempts to return the hand to a previously observed solution (perhaps utilizing motor memory (Ganesh et al., 2010)) allows for sufficient error information to be collected for an internal model update (Kawato and Gomi, 1992). Once a reliable, local internal model is formed, and the ability to reach the target is restored, the CNS can then turn its efforts towards energy optimization.

In conclusion, our work addresses the intrinsic objective of a point-to-point reaching movement. While there are notable arguments made for the parsimonious case of only terminal accuracy and minimal motor effort, we suggest that robustness, especially against model uncertainly, should additionally be considered essential for reaching. A kinematic invariant, independent of the underlying dynamics, helps the CNS mitigate the risks of uncertainty while enabling the eventual learning of a more efficient control solution.
Acknowledgments

We thank Gary Liaw for his contributions during initial pilot studies of this experiment and Toshinori Yoshioka for implementation and operation of the experimental equipment.

Grants

This research was supported in part by National Science Foundation grants ECS-0326095, IIS-0535282, IIS-1017134, CNS-0619937, IIS-0917318, CBET-0922784, EECS-0926052, CNS-0960061, the DARPA program on Advanced Robotic Manipulation, the Army Research Office, the Okawa Foundation, and the Strategic Research Program for Brain Sciences “Brain Machine Interface Development” by the Ministry of Education, Culture, Sports, Science and Technology of Japan.
References


Figure Legends

Figure 1. The ATR Planar Force Manipulandum (PFM).

Figure 2. Our acceleration force field shown as arrows for a minimum jerk reaching movement. The force field, which only acts in the $x$ direction, depends on $y$ acceleration.

Figure 3. The prediction made by stochastic optimal control under a rightward viscous curl field. Each line plots the average of 50 simulations. Using a non-directional cost function (SIM.nd), the optimal action (shown by the thickest line) is to move with a significant leftward curvature, first working against the field while velocity is low, then subsequently allowing the field to push the hand back towards the target. While under the directional cost function (SIM.d), the optimal action is to restore the trajectory towards baseline. Before Effects (medium thickness line) show the action before learning occurs ($\alpha = 0.00$), while thin lines indicate incremental stages of learning ($\alpha = 0.25, 0.50, 0.75$).

Figure 4. Top row shows trajectories and simulations from Experiment A. Bottom row is from Experiment B. In DATA plots (rightmost), each line is the average trajectory of all subjects (A.DATA: 6 subjects; B.DATA: 10 subjects). Baseline trajectories (long dashed lines) are the average of the last 15 trials in null field training. Final Force (thickest solid line) are the average of the last 15 trials in force field training. Before Effects (grey solid line, A.DATA only) averages the 10 catch trials following null field training. After Effects (short dashed lines) averages the 10 catch trials following force field training. Thin solid lines show incremental stages of learning, the statistics of which are reported in Table 1 (for A.DATA) and Table 2 (for B.DATA). The leftmost column shows simulations of the stochastic optimal controller using the non-directional cost function (SIM.nd). The center column uses directional cost (SIM.d). Each SIM line is the average of 50 simulations. Under either cost function, preadaptation behavior in the force field (Before Effects, $\alpha = 0.0, \beta = 1.0$) is a rightward curved trajectory that misses the target to the left. Incremental learning is represented by thin lines (with $\alpha = 0.25, 0.50, 0.75$) until complete learning (Final Force, $\alpha = 1.0$). In the non-directional case (A.SIM.nd), the trajectory adapts outward, with increasing curvature. However, in the directional case (A.SIM.d), adaptation occurs inward with decreasing curvature. Similarly, a gradual increase in field strength predicts gradually increasing curvature in the non-directional case (B.SIM.nd) and little change in the directional case (B.SIM.nd). Thin lines in B.SIM are: $\beta = 0.25, 0.50, 0.75$, with thickest line $\beta = 1.0$ (and perfect estimation is assumed, $\alpha = 1.0$, as field strength is increased).
Figure 5. Top: Averaged trial-by-trial learning curve during Block 3 of Experiment A (the first block of force field training). Bottom: Averaged trial-by-trial learning curve during Block 2 of Experiment B, which includes the ramp stage of gradually increasing the force field gain from zero (indicated by rising gray dashed line). In this case, the trial sequences of each subject are aligned by the start of the ramp stage. The $y$-axis in the plots is the average distance from the line connecting the start of the trial to the target. The baseline value is also shown with a dashed line. For selected trials, error bars indicate plus and minus one standard deviation. The thick line is a 15 point moving average.

Figure 6. DATA: Forces in the $x$ direction (the direction of the force field perturbation) as measured by the force sensor at the handle of the manipulandum. DATA (averaged) shows the average among all 6 subjects of Experiment A. DATA (individual) shows individual trials from a single subject of Experiment A. Baseline trajectories (dashed lines) are the last 15 trials in null field training. Final Force (thick solid lines) are the last 15 trials in force field training. Before Effects (thin solid lines) are the 10 catch trials following null field training. While the baseline force profile has a rightward (positive) force (likely due to subjects compensating for manipulandum dynamics), the force field disturbance elicits a strong corrective force in the leftward (negative) direction in the early part of the movement. SIM: Output force, in the $x$ direction, of the simulated optimal controller under the full strength force field (Experiment A). The model with non-directional cost function initially pushes in the same direction as the disturbance, while the directional model applies similar resistive forces as demonstrated by the subjects.

Figure 7. The position and velocity gains of the optimal controller, in the $x$ direction (same direction as the force perturbation) before and after adding a directional constraint to the cost function. Adding the directional constraint effectively increases stiffness and damping during the early part of the movement.

Figure 8. Sensitivity of the controller with respect to unknown changes in arm mass. In each case, the optimal controller is computed assuming the nominal mass value, but tested on 50%, 75%, 100%, 125%, and 150% of the mass value. SIM.nd uses the non-directional cost function, and SIM.d uses directional cost.
Tables

Table 1. Statistics for Experiment A

<table>
<thead>
<tr>
<th>Trajectory</th>
<th>Mean (m)</th>
<th>Stdev (m)</th>
<th>P</th>
<th>Sig.</th>
<th>Success Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>-0.0020</td>
<td>0.0018</td>
<td></td>
<td></td>
<td>58.9%</td>
</tr>
<tr>
<td>Before Effects</td>
<td>0.0128</td>
<td>0.0028</td>
<td>3.0 × 10^{-6}</td>
<td>**</td>
<td>20.0%</td>
</tr>
<tr>
<td>Force Trials [1-10]</td>
<td>0.0086</td>
<td>0.0030</td>
<td>6.6 × 10^{-5}</td>
<td>**</td>
<td>25.0%</td>
</tr>
<tr>
<td>Force Trials [11-20]</td>
<td>0.0050</td>
<td>0.0036</td>
<td>3.6 × 10^{-3}</td>
<td>**</td>
<td>25.0%</td>
</tr>
<tr>
<td>Force Trials [21-30]</td>
<td>0.0038</td>
<td>0.0033</td>
<td>6.5 × 10^{-3}</td>
<td>**</td>
<td>36.7%</td>
</tr>
<tr>
<td>Force Trials [31-40]</td>
<td>7.7 × 10^{-4}</td>
<td>0.0050</td>
<td>0.2500</td>
<td></td>
<td>61.7%</td>
</tr>
<tr>
<td>Final Force</td>
<td>0.0032</td>
<td>0.0054</td>
<td>0.0694</td>
<td></td>
<td>65.6%</td>
</tr>
<tr>
<td>After Effects</td>
<td>-0.0131</td>
<td>0.0043</td>
<td>7.5 × 10^{-4}</td>
<td>**</td>
<td>16.7%</td>
</tr>
</tbody>
</table>

The trajectories Baseline, Before Effects, Final Force, and After Effects are as defined and plotted in Figure 4 (A.DATA). Force Trials represent bins of 10 of trials each from Block 3 (the first block of force field training), and plotted as thin lines in Figure 4 (A.DATA). Mean and Stdev report the mean trajectory deviation and standard deviation (respectively) of each trajectory for the 6 subjects of experiment A. P and Sig. report the significance from baseline (Paired t-test, df=5, **=P<0.01). Success rates for each trajectory are also shown.

Table 2. Statistics for Experiment B

<table>
<thead>
<tr>
<th>Trajectory</th>
<th>Mean (m)</th>
<th>Stdev (m)</th>
<th>P</th>
<th>Sig.</th>
<th>Success Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>-0.0021</td>
<td>0.0036</td>
<td></td>
<td></td>
<td>58.0%</td>
</tr>
<tr>
<td>Ramp Trials [1-10]</td>
<td>0.0011</td>
<td>0.0047</td>
<td>0.1125</td>
<td></td>
<td>59.0%</td>
</tr>
<tr>
<td>Ramp Trials [11-20]</td>
<td>0.0025</td>
<td>0.0058</td>
<td>0.0511</td>
<td></td>
<td>34.0%</td>
</tr>
<tr>
<td>Force Trials [1-10]</td>
<td>0.0019</td>
<td>0.0062</td>
<td>0.1015</td>
<td></td>
<td>23.0%</td>
</tr>
<tr>
<td>Force Trials [11-20]</td>
<td>4.5 × 10^{-4}</td>
<td>0.0056</td>
<td>0.2498</td>
<td></td>
<td>34.0%</td>
</tr>
<tr>
<td>Final Force</td>
<td>0.0015</td>
<td>0.0048</td>
<td>0.0768</td>
<td></td>
<td>56.0%</td>
</tr>
<tr>
<td>After Effects</td>
<td>-0.0117</td>
<td>0.0017</td>
<td>4.0 × 10^{-6}</td>
<td>**</td>
<td>24.0%</td>
</tr>
</tbody>
</table>

The trajectories Baseline, Before Effects, Final Force, and After Effects are as defined and plotted in Figure 4 (B.DATA). Ramp [1-10] and Ramp [11-20] (plotted as thin lines in B.DATA) are the first and last 10 trials, respectively, during the ramp stage where the force field gain gradually increases from trial to trial. Post Ramp [1-10] and Post Ramp [11-20] (also thin lines in B.DATA), are the first 10 and next 10 trials immediately following the ramp stage (with force field at full strength). Mean and Stdev report the mean trajectory error and standard deviation (respectively) for the 10 subjects of experiment B. P and Sig. report the significance from baseline (Paired t-test, df=9, **=P<0.01). Success rates for each trajectory are also shown.
DATA (averaged)

DATA (individual)

SIM