Impedance Control as an Emergent Mechanism from Minimising Uncertainty

Djordje Mitrovic, Stefan Klanke, Rieko Osu, Mitsuo Kawato, Sethu Vijayakumar

February 20, 2009

Abstract

Efficient human motor control is characterised by an extensive use of joint impedance modulation, which is achieved by co-contracting antagonistic muscle pairs in a way that is beneficial to the specific task. While there is much experimental evidence available that the Central Nervous System (CNS) employs such an impedance control strategy only few computational models of impedance control have been proposed so far. In this paper, we study the computational aspects of the generation of joint impedance control in human arm reaching tasks and we develop a new model of joint impedance control for antagonistic systems. We formulate an actor's goal of arm reaching by optimising a cost function that accounts for maximal positional accuracy and minimal energy expenditure. To account for shortcomings of previously presented optimal control models, that fail to model impedance control, we employ the concept of learned internal dynamics models in conjunction with a stochastic arm simulation model, that exhibits realistic signal dependent noise-impedance characteristics. When using this stochastic arm model for dynamics learning with a locally weighted learning algorithm, the produced noise or kinematic variability reflects the prediction uncertainty on the level of the internal model. Introducing this information into the stochastic Optimal Feedback Control (OFC) theory reveals that impedance control naturally emerges from an optimisation process that minimises for model prediction uncertainties, along with energy and accuracy demands. We evaluate our method in single-joint simulations under static reaching conditions as well as in adaptation scenarios. The results show that our model is able to explain many well-known impedance control patterns from the literature, which supports our impedance control model as a viable approach of how the CNS could modulate joint impedance.

1 Introduction

Humans and other biological systems have excellent capabilities in performing fast and complicated control tasks in spite of large sensorimotor delays, internal noise or external perturbations. By co-activating antagonistic muscle pairs, the CNS manages to change the mechanical properties (i.e., joint impedance) of limbs in response to specific task requirements; this is commonly referred to as *impedance control* (Hogan, 1984). A significant benefit of modulating the joint impedance is that the changes apply instantaneously to the system. Impedance control has been explained as an effective strategy of the CNS to cope with kinematic variability due to neuromuscular noise and environmental disturbances. Understanding how the CNS realises impedance control is of central interest in biological motor control as well as in the control theory of artificial systems.

The role of adaptive joint stiffness in humans has been investigated in static (e.g., (Perreault et al., 2001; Selen et al., 2005)) and dynamic tasks (e.g., (Burdet et al., 2001; Osu et al., 2004)). Studies in single and multi joint limb reaching movements revealed that stiffness is increased with faster movements (Suzuki et al., 2001) as well as with higher positional accuracy demands (i.e.smaller reaching targets) (Gribble et al., 2003). Other work has investigated impedance modulation during adaptation towards external force fields and (Burdet et al., 2001) showed that

human subjects improved reaching performance when faced with unstable dynamics[1] by learning optimal mechanical impedance of their arm. This experiment showed that subjects are able to predictively control the magnitude, shape, and orientation of the endpoint stiffness without varying endpoint force. Therefore the joint impedance can be understood as an additional degree of freedom, which can be controlled independently from the joint torque by co-contracting antagonistic muscles (Osu and Gomi, 1999). Recently (Franklin et al., 2008) have presented a computational motor learning model, which proposes that the CNS optimises impedance along with accuracy and energy-efficiency. Adaptation patterns in human subjects showed that co-contraction decreases over the course of practice; these learning effects were observed to be stronger in stable force fields (i.e., velocity-dependent) compared to unstable force fields (i.e., divergent), which suggests that impedance control is linked to the learning process with internal dynamics models and that the CNS uses impedance control to increase task accuracy in early learning stages, when the internal model is not fully formed yet.

While behavioral studies have emphasized the importance of impedance control in the CNS, relatively few computational models have been proposed (Tee et al., 2004; Burdet et al., 2006; Franklin et al., 2008). This paper is concerned with impedance control during arm reaching tasks and we present a new model which can predict many well-documented co-activation patterns observed in humans. Our model is formalised within the framework of *stochastic Optimal Feedback Control (OFC)* (Todorov and Jordan, 2002; Todorov, 2005), which has been very successful in explaining human reaching movements as a minimisation process of motor energy and kinematic end-point error (Liu and Todorov, 2007). OFC presents itself as a powerful theory for interpreting biological motor control (Scott, 2004), since it unifies motor costs, expected rewards, internal models, noise and sensory feedback into a coherent mathematical framework (Shadmehr and Krakauer, 2008). For the study of biological systems it is furthermore well justified a priori because optimal control often is interpreted as a results of *natural optimisation* (i.e. evolution, learning, adaptation).

Most OFC models assumed perfect knowledge of the system dynamics, given in closed analytic form based on the equations of motion. Such computations usually make simplifying rigid body assumptions and for complex systems the required model parameters may be unknown, hard to estimate or even subject to changes, which additionally complicates to model "unpredictable" perturbations. In a biological context it remains unclear how the analytic dynamics equations translate into a plausible neural representation of the internal dynamics model. In order to overcome these limitations we postulate that the dynamics model evolve from a motor learning process, in which the dynamics model is modified regularly from previously experienced sensorimotor inputs. Everyday experience shows that humans are able to learn from changing environments and a vast number of studies (for a review see (Davidson and Wolpert, 2005)) suggest that the motor system forms an internal forward dynamics model of its arm and the perturbations applied to the hand, visual scene or the target. This internal model helps to compensate for delays, uncertainty of sensory feedback, and environmental changes in a predictive fashion (Wolpert et al., 1995; Kawato, 1999). A supervised learning based formalism therefore seems a plausible neural representation of the system dynamics and incorporating this learned internal model within the OFC framework allows for a modelling of adaptation processes. By updating the internal model with arm data produced during control, this *OFC model with learned dynamics (OFC-LD)* can explain human trial-to-trial adaptation patterns towards external force fields (FF) (Mitrovic et al., 2008a).

Given the stochastic nature of the sensorimotor system (Faisal et al., 2008), motor control theories, in order to be efficient, must be able to account for the resulting effects of *signal dependent noise (SDN)*. Early work on stochastic optimal control was based on the assumption that noise limits the information capacity of the motor system, which revealed a speed-accuracy trade-off in reaching movements, known as *Fitts' Law* (Fitts, 1954). More recent work in the

[1] Created using a divergent force field.

framework of stochastic optimal control (Harris and Wolpert, 1998) formulated reaching tasks as an optimal trade-off between task achievement and a minimisation of the corruptive effects of SDN. These models have been successful in reproducing Fitts' law. Extensions described for example obstacle avoidance (Hamilton and Wolpert, 2002) and step tracking wrist movements (Haruno and Wolpert, 2005). When noisy feedback is taken into consideration, (Todorov and Jordan, 2002) showed that the *minimum intervention principle* and *motor synergies* emerge naturally within the framework of stochastic OFC framework. Even though those models take into account signal dependent noise they essentially ignore the impedance to noise characteristics of the musculoskeletal system (Section 2) and therefore are only concerned with finding the lowest muscle activation possible for achieving a specific task under *naive SDN*. Experimental data however suggests that the CNS "sacrifices" energetic costs of muscles to reach stability in form of higher joint impedance under certain conditions. While similar assumptions have been stated previously (Gribble et al., 2003; Osu et al., 2004) no conclusive OFC model has been presented so far. To account for that major drawback we make the assumption that our internal model has been learned from a system that exhibits *realistic SDN* and kinematic variability (as it is the case in humans). We use a local learning method that provides us with statistical information about the motor variability in the form of heteroscedastic prediction variances, which can be interpreted as a representation of the certainty of the internal model predictions. We then can formulate a *minimum-uncertainty* optimal control model that introduces this stochastic information into OFC. This has the beneficial effect that our model favours co-contraction in order to reduce the negative effects of SDN and at the same time tries to minimise energy cost and endpoint reaching error. For finding the stochastic optimal feedback control law we employ computational methods that iteratively compute an optimal trajectory together with a locally valid feedback law and therefore avoid the *curse of dimensionality* global OFC methods typically suffer from.

2 An Antagonistic Arm Model for Impedance Control

We want to study impedance control in planar single-joint reaching movements under different task conditions such as initial or final position, different speeds and adaptation towards external forces. The single joint reaching paradigm is a well accepted experimental paradigm to investigate simple human reaching behaviour (Osu et al., 2004) and the arm model presented here mimics planar rotation about the elbow joint using two elbow muscles.

The dynamics of the arm is in part based on standard equations of motion. The joint torques $\boldsymbol{\tau}$ are given by

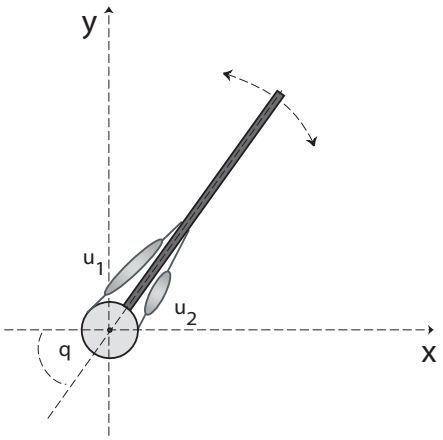
$$\boldsymbol{\tau} = \mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}}, \tag{1}$$

where $\mathbf{q}$ and $\dot{\mathbf{q}}$ are the joint angles and velocities, respectively; $\mathbf{M}(\mathbf{q})$ is the symmetric joint space inertia matrix, which in the one joint planar case is a constant $\mathbf{M}(\mathbf{q}) = \mathbf{M}$. The Coriolis and centripetal forces are accounted for by $\mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})$. The joint torques produced by a muscle are a function of its moment arms, the muscle tension, and the muscle activation dynamics. To compute effective torques from the muscle commands $\mathbf{u}$ the corresponding transfer function is given by

$$\boldsymbol{\tau}(\mathbf{q}, \dot{\mathbf{q}}, \mathbf{u}) = -\mathbf{A}(\mathbf{q})^T \mathbf{t}(\mathbf{l}, \dot{\mathbf{l}}, \mathbf{u}), \tag{2}$$

where $\mathbf{A}(\mathbf{q})$ represents the moment arm. For simplicity, we assume $\mathbf{A}(\mathbf{q}) = \mathbf{A}$ to be constant and independent of the joint angles $\mathbf{q}$. The values of $\mathbf{u}$ are assumed to be non-negative within the range $[0, 1]$. The muscle lengths $\mathbf{l}$ depend on the joint angles $\mathbf{q}$ through the affine relationship $\mathbf{l} = \mathbf{l}_m - \mathbf{A}\mathbf{q}$, which also implies $\dot{\mathbf{l}} = -\mathbf{A}\dot{\mathbf{q}}$. The term $\mathbf{t}(\mathbf{l}, \dot{\mathbf{l}}, \mathbf{u})$ in (2) denotes the muscle tension, for which we follow a spring-damper model defined as

$$\mathbf{t}(\mathbf{l}, \dot{\mathbf{l}}, \mathbf{u}) = \mathbf{k}(\mathbf{u})(\mathbf{l}_r(\mathbf{u}) - \mathbf{l}) - \mathbf{b}(\mathbf{u})\dot{\mathbf{l}}. \tag{3}$$



Arm parameters	
Link weight [kg]	$m = 1.59$
Link length [m]	$l = 0.35$
Center of gravity [m]	$L_g = 0.18$
Moment of inertia [kg $\cdot$ m^2]	$I = 0.0477$
Moment arms [cm]	$\mathbf{A} = [2.5\ 2.5]^T$

Muscle parameters	
Elasticity $[N/m]$	$k = 1621.6$
Intrinsic elasticity $[N/m]$	$k_o = 810.8$
Viscosity $[N \cdot s/m]$	$b = 108.1$
Intrinsic viscosity $[N \cdot s/m]$	$b_0 = 54.1$
Rest length constant	$r = 2.182$
Muscle length at rest position [cm] ($\mathbf{q_0} = \pi/2$)	$l_0 = 2.637$

Figure 1: Left: Human elbow model with two muscles. Right: Used arm and muscle parameters (Adapted from (Katayama and Kawato, 1993)). Flexor and extensor muscles are modelled with identical parameters.

Here, $\mathbf{k}(\mathbf{u})$, $\mathbf{b}(\mathbf{u})$, and $\mathbf{l}_r(\mathbf{a})$ denote the muscle stiffness, the muscle viscosity and the muscle rest length, respectively. Each of these terms depends linearly on the muscle signal $\mathbf{u}$, as given by

$$\mathbf{k}(\mathbf{u}) = diag(\mathbf{k}_0 + k\mathbf{u}), \qquad \mathbf{b}(\mathbf{u}) = diag(\mathbf{b}_0 + b\mathbf{u}), \qquad \mathbf{l}_r(\mathbf{u}) = \mathbf{l}_0 + r\mathbf{u}. \tag{4}$$

The elasticity coefficient k, the viscosity coefficient b, and the constant r are given from the muscle model. The same holds true for $\mathbf{k}_0$, $\mathbf{b}_0$, and $\mathbf{l}_0$, which are the intrinsic elasticity, viscosity and rest length for $\mathbf{u} = \mathbf{0}$, respectively. Figure 2 depicts the arm model and its parameters, which have been adapted from (Katayama and Kawato, 1993).

With the presented model we can formulate the forward dynamics as

$$\ddot{\mathbf{q}} = \mathbf{M}^{-1}(\boldsymbol{\tau}(\mathbf{q}, \dot{\mathbf{q}}, \mathbf{u}) - \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}}). \tag{5}$$

To model the stochastic nature of neuromuscular signals many proposed models simply contaminate the neural inputs $\mathbf{u}$ with multiplicative noise, for example with a standard deviation of 20% of the signal's magnitude ($\sigma_u = 0.2$) (Li, 2006). Such an approach models just naive control-dependent noise but cannot account for the complex interplay of neuromuscular noise, modified joint impedance and kinematic variability.

Kinematic variability in human motion originates from a combination of the effects of variability of muscle forces and from environmental perturbations. In the presence of an external perturbation, it appears obvious that increased joint stiffness will stabilise the motion towards a planned trajectory (Burdet et al., 2006). Internal force fluctuations are inevitable due to the stochastic nature of neuromuscular processes and the causalities here are less intuitive: Muscular force fluctuations (Jones et al., 2002) as well as joint impedance (Osu and Gomi, 1999) increase monotonically with the level of muscle co-activation leading to the paradoxical situation that muscles are the source of fluctuation and at the same time the means to suppress its effect by increasing joint impedance (Selen et al., 2005). Details about the sources of neuromotor noise and its signal dependency are discussed in (Selen, 2007; Faisal et al., 2008).

In order to describe a realistic antagonistic simulation model that produces reduced kinematic variability despite increased noise levels in the muscles, the model must most importantly produce appropriate force variability. In isometric[2] contraction studies in simulation (Selen

[2]Here defined as that type of contraction where the joint angle is held constant. In contrast, *isotonic* contraction induces joint angle motion.

et al., 2005) showed that standard Hill-type muscle models, similar to our spring-damper-system, fail to produce appropriate increase of force variability as observed in humans. The authors present a motor unit pool model of parallel Hill-type motor units which could achieve the desired motor variability. Therefore higher co-contraction can be understood as a low-pass filter to the kinematic variability, showing that higher joint impedance is in principle an effective strategy to meet higher accuracy demands, given that the muscles forces are modelled correctly.

So far studies on muscle force variability (in simulation as well as with human subjects) have investigated isometric contractions only, whereas we are primarily interested in the computational nature of impedance control during reaching movements (i.e., isotonic contractions). As an alternative we therefore propose to increase the realism of our arm model by imposing the kinematic variability based on physiological observations that variability increases monotonically with higher muscle activations, while the variability is reduced for more highly co-contracted activation patterns. We further make the reasonable assumption that isotonic contraction causes larger variability than pure isometric contraction. In reality, at very high levels of co-contraction synchronisation effects may occur, which become visible as tremor of the arm (Selen, 2007). We will ignore such extreme conditions in our model. Based on the stated assumptions we can formulate the kinematic variability depending on muscle co-activation of antagonistic muscle pairs as a noise process $\xi(u)$. The regular muscle tension calculation then can be extended to be

$$t_i^{ext}(l_i, \dot{l}_i, u_i) = t_i(l_i, \dot{l}_i, u_i) + \xi(u_i). \tag{6}$$

The variability in muscle tensions depending on antagonistic muscle activations can be modelled as

$$\xi(u_i) \sim \mathcal{N}(0, \sigma_{isotonic}|u_i - u_i'|^n + \sigma_{isometric}|u_i + u_i'|^m). \tag{7}$$

In eq. 7 indices u_i and u_i' indicate antogonistic muscle pairs. The first term (of the distribution's standard deviation) weighted with a scalar $\sigma_{isotonic}$ accounts for increasing variability in isotonic muscle contraction, while the second term accounts for the amount of variability for co-contracted muscles. Parameters $n, m \in \Re$ act as additional curvature-shaping parameters that define the shape of the monotonic increase of the SDN. The resulting contraction variability

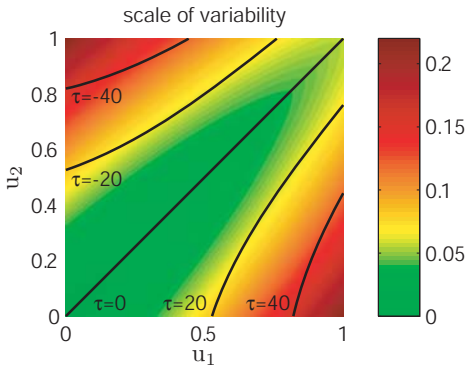


Figure 2: Induced function of variability as defined in eq. 7 with parameters $\sigma_{isotonic} = 0.2$, $\sigma_{isometric} = 0.02$, $n = 1.5$, $m = 1.5$. The black lines indicate the muscle activations that produice equal joint torques for $\boldsymbol{\tau} = [-40, -20, 0, 20, 40]$.

relationship produces plausible muscle tension characteristics without introducing highly complex parameters into the arm model. Figure 2 shows the produced joint torques of our model for all combinations of u_1 and u_2 for constant joint angles ($\mathbf{q} = \frac{\pi}{2}$) and joint velocities ($\dot{\mathbf{q}} = 0$). From the right plot we can see that co-contraction reduces the variability in the joint torques.

To see how the induced variability in the muscle tensions translates into the joint acceleration we can formulate the forward dynamics including the variability as

$$\ddot{\mathbf{q}}^{ext} = \mathbf{M}^{-1}(\boldsymbol{\tau}^{ext}(\mathbf{q}, \dot{\mathbf{q}}, \mathbf{u}) - \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}}). \tag{8}$$

With

$$\boldsymbol{\tau}^{ext}(\mathbf{q}, \dot{\mathbf{q}}, \mathbf{u}) = -\mathbf{A}^T \mathbf{t}^{ext}(\mathbf{l}, \dot{\mathbf{l}}, \mathbf{u}) = -\mathbf{A}^T \mathbf{t}(\mathbf{l}, \dot{\mathbf{l}}, \mathbf{u}) - \mathbf{A}^T \boldsymbol{\xi}(\mathbf{u}) \tag{9}$$

we get an equation of motion including a noise term

$$\ddot{\mathbf{q}}^{ext} = \mathbf{M}^{-1}(\boldsymbol{\tau}(\mathbf{q}, \dot{\mathbf{q}}, \mathbf{u}) - \mathbf{A}^T \boldsymbol{\xi}(\mathbf{u}) - \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}}). \tag{10}$$

Multiplying all terms and rearranging them leads to following extended forward dynamics equation

$$\ddot{\mathbf{q}}^{ext} = \ddot{\mathbf{q}} - \mathbf{M}^{-1}\mathbf{A}^T \boldsymbol{\xi}(\mathbf{u}), \tag{11}$$

which is separated into a deterministic component $\mathbf{f}(\mathbf{q}, \dot{\mathbf{q}}, \mathbf{u}) = \ddot{\mathbf{q}}$ and a stochastic part $\mathbf{F}(\mathbf{u}) = \mathbf{M}^{-1}\mathbf{A}^T \boldsymbol{\xi}(\mathbf{u})$. One should note that the stochastic component in our case is only dependent on the muscle signals $\mathbf{u}$, because the matrices $\mathbf{A}$ and $\mathbf{M}$ are independent of the arm states. However this can be easily extended for more complex arm models with multiple links or state-dependent moment arms.

As shown, the extended noise corresponds to an additional noise term in the joint accelerations which is directly linked to kinematic variability through integration over time. Therefore for the rest of this paper we refer to *noise* as an equivalent to *kinematic variability*. Even though the presented dynamics and noise model is a rather simplistic representation for a real human limb, it suffices in that the nonlinear antagonistic formulation paired with the biophysically plausible tension noise allows us to model realistic impedance control that stabilises the plant, both in the presence of external perturbations and due to internal fluctuations.

3 Finding the Optimal Control Law

Let $\mathbf{x}(t)$ denote the state of the arm model and $\mathbf{u}(t)$ the applied control signal at time t. The state consists of the joint angles $\mathbf{q}$ and velocities $\dot{\mathbf{q}}$. The control signals correspond to the neural control input denoted by $\mathbf{u}$. If the system would be deterministic, we could express its dynamics as $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u})$, whereas in the presence of noise we write the dynamics as a stochastic differential equation

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, \mathbf{u})dt + \mathbf{F}(\mathbf{x}, \mathbf{u})d\boldsymbol{\omega}. \tag{12}$$

Here, $d\boldsymbol{\omega}$ is assumed to be Brownian motion noise, which is transformed by a possibly state- and control-dependent matrix $\mathbf{F}(\mathbf{x}, \mathbf{u})$. The optimal control problem can be stated as follows: Given an initial state $\mathbf{x}_0$ at time $t = 0$, we seek a control sequence $\mathbf{u}(t)$ such that the system's state is $\mathbf{x}^*$ at end-time $t = T$. Optimal control theory approaches the problem by first specifying a cost function which is composed of (i) some evaluation $h(\mathbf{x}(T))$ of the final state, usually penalising deviations from the desired state $\mathbf{x}^*$, and (ii) the accumulated cost $c(t, \mathbf{x}, \mathbf{u})$ of sending a control signal $\mathbf{u}$ at time t in state $\mathbf{x}$, typically penalising large motor commands. Introducing a policy $\boldsymbol{\pi}(t, \mathbf{x})$ for selecting $\mathbf{u}(t)$, we can write the expected cost of following that policy from time t as (Todorov and Li, 2005)

$$v^{\boldsymbol{\pi}}(t, \mathbf{x}(t)) = \Big\langle h(\mathbf{x}(T)) + \int_t^T c(s, \mathbf{x}(s), \boldsymbol{\pi}(s, \mathbf{x}(s)))ds \Big\rangle. \tag{13}$$

One then aims to find the policy $\boldsymbol{\pi}$ that minimises the total expected cost $v^{\boldsymbol{\pi}}(0, \mathbf{x}_0)$. Thus, in contrast to classical control, calculation of the trajectory (planning) and the control signal (execution) is not separated anymore, and for example, redundancy can actually be exploited in order to decrease the cost. The dynamics $\mathbf{f}$ of our arm model is highly non-linear in $\mathbf{x}$ and $\mathbf{u}$ and it does not fit into the Linear Quadratic framework (Stengel, 1994), which motivates the use of approximative OFC methods.

Differential dynamic programming (DDP) (Jacobson and Mayne, 1970) is a well-known successive approximation technique for solving nonlinear dynamic optimisation problems. This

method uses second order approximations of the system dynamics to perform dynamic programming in the neighbourhood of a nominal trajectory. A more recent algorithm is the iterative Linear Quadratic Regulator (ILQR) (Li and Todorov, 2004). This algorithm uses iterative linearisation of the nonlinear dynamics around the nominal trajectory, and solves a locally valid LQR problem to iteratively improve the trajectory. However, this method is still deterministic and cannot deal with control constraints or non-quadratic cost functions. A recent extension to ILQR, the *iterative Linear Quadratic Gaussian (ILQG)* framework (Todorov and Li, 2005), allows to model nondeterministic dynamics by incorporating a Gaussian noise model. Furthermore it supports control constraints like non-negative muscle activations or upper control boundaries. The ILQG framework showed to be computationally significantly more efficient than DDP (Li and Todorov, 2004). It has also been previously tested on biological motion systems and therefore is the favourite approach for calculating the optimal control law. The ILQG algorithm is outlined in Appendix A and for implementation details we refer the reader to (Todorov and Li, 2005).

We are studying reaching movements of a finite time horizon of length $T = k\Delta t$ seconds. Typical values for a 0.5 seconds simulation are $k = 50$ discretisation steps with simulation rate of $\Delta t = 0.01$. For a typical reaching task we define a cost function of the form

$$v = w_p \mid \mathbf{q}_T - \mathbf{q}_{tar} \mid^2 + w_v \mid \dot{\mathbf{q}}_T \mid^2 + w_e \sum_{k=0}^{T} \mid \mathbf{u}(k) \mid^2 \Delta t. \tag{14}$$

The first term penalises reaches away from the target joint angle $\mathbf{q}_{tar}$, the second term forces a zero velocity at the end time T, and the third term penalises large muscle commands (i.e., minimises energy consumption) during reaching. The factors w_p, w_v, and w_e weight the importance of each component. We have found that ILQG works well on reaching task using our arm model. For this example the arm started at zero velocity and start position $\mathbf{q}_0 = \frac{2 \cdot \pi}{3}$ and aimed to reach $\mathbf{q}_{tar} = \frac{\pi}{2}$ in $T = 500ms$. Figure 3 shows a typical motion generated by using ILQG in a deterministic scenario. It generates the characteristic bell-shaped velocity profiles and a muscle activation pattern where the first peak accelerates the limb and the second peak causes deacceleration and stopping at the target. As expected the muscle activations show minimal co-contraction during reaching due to the imposed minimum energy performance criterion. In this paper we define co-contraction as the minimum of two antagonistic muscle pairs $min(u_1, u_2)$ (Thoroughman and Shadmehr, 1999).

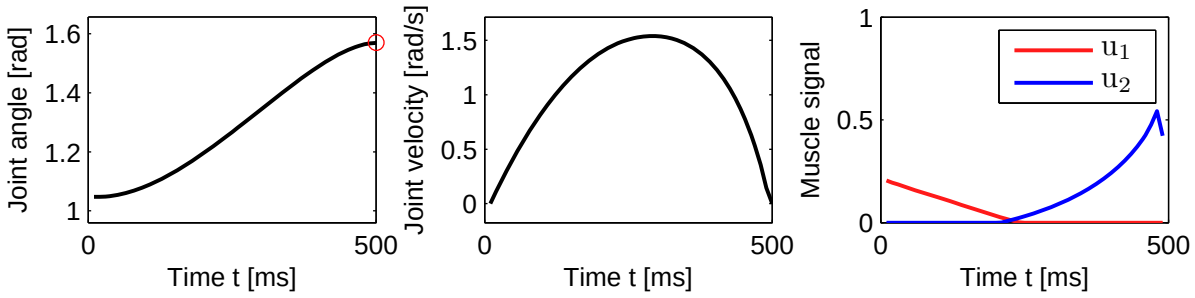
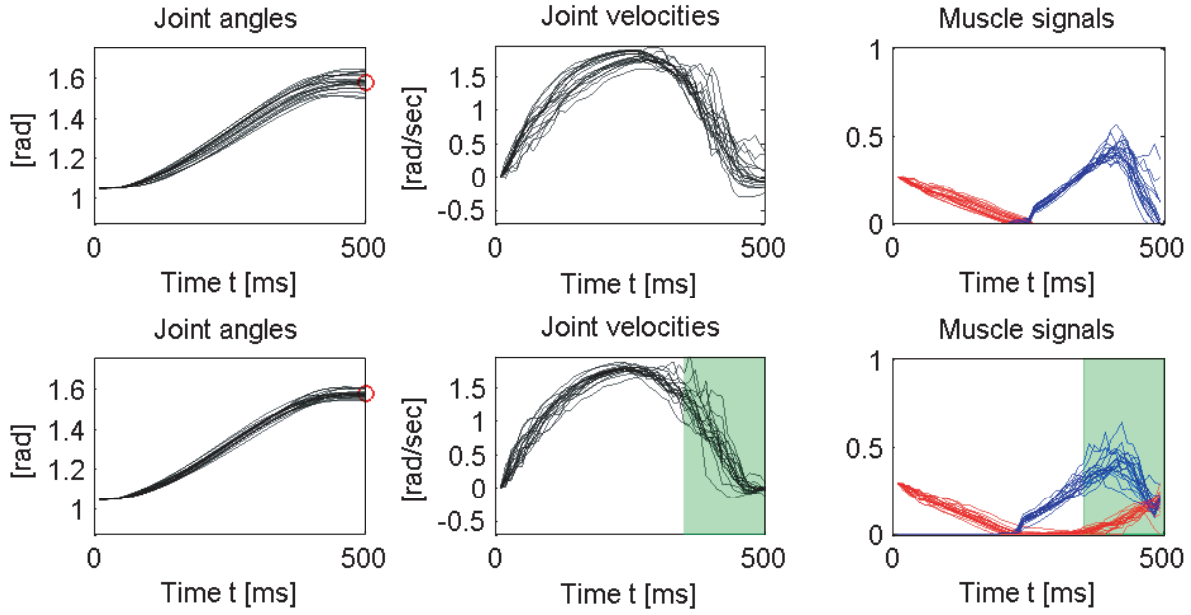


Figure 3: Single joint optimal reaching using ILQG. Left: Joint angle trajectory where the red circle indicates the target to reach at 500 milliseconds. Middle: Characteristic bell-shaped joint angle velocity profile. Right: Muscle signals which show virtually no co-activation due to the minimum energy performance criterion.

4 Uncertainty Driven Impedance Control

The previous example modelled the deterministic case where we assumed that the plant is noise free and matrix $\mathbf{F}(\mathbf{x}, \mathbf{u})$ to have zero entries. Here we want to elucidate upon the stochastic scenario where the plant suffers from realistic SDN as presented in equation (7). Following the stochastic OFC formulation we can set the stochastic component $\mathbf{F}(\mathbf{x}, \mathbf{u}) = \mathbf{F}(\mathbf{u})$ as defined in eq. (11). For the moment we do not model any state-dependencies in the noise term. By introducing stochastic information, ILQG will now perform an optimisation that takes into account the control-dependent "shaped" noise of the system. This leads to optimal control solutions that minimise the negative effects of the noise, which by definition of $\mathbf{F}(\mathbf{u})$ should increase in isometric contraction (i.e., co-contraction) and therefore increase the joint impedance. To illustrate this effect we repeat the ILQG reaching experiment from the previous section. However now we use the noisy arm model and we control it in closed loop control scheme, using ILQG's feedback control law given by the matrix $\mathbf{L}$ (see Appendix A) to correct the plant from the effects of the SDN.



	Joint angle error [rad]	Joint velocity error [rad/s]	Energy consumption
Naive SDN	**0.03026** ± 0.02197	**0.12192** ± 0.12543	**0.09252** ± 0.00813
Extended SDN	**0.01797** ± 0.01499	**0.02950** ± 0.03547	**0.11464** ± 0.00533

Figure 4: Comparison of the performance of stochastic ILQG with naive SDN (first row plots) and extended SDN (second row plots). For clearer visualisation only the first 20 trajectories are plotted. The shaded green area indicates the region in which the extended noise solution exhibits increasing co-contraction. The table quantifies the results (mean ± standard deviation). First table row: average joint angle error (absolute values) at final time T. Second table row: Joint angle velocity (absolute values) at time T. Third table row: integrated muscle commands (of both muscles) over trials. The extended SDN outperforms the reaching performance of the naive SDN case, with the price of a higher energy consumption.

Figure 4 compares the reaching with stochastic ILQG with naive SDN and with extended SDN. We performed 50 reaching movements (only 20 trajectories plotted) with the two different noisy plants. The table within Figure 4 quantifies the results and reveals that the extended SDN performs significantly better in terms of end point accuracy and end point velocity. Even though the naive SDN solution tries to reduce the negative effects of the noise, by applying very low

muscle commands at the end of the motion, it still fails to stabilise the plant towards the end of the motion. In contrast the extended SDN co-contracts towards the end of the motion in order to reduce the negative effects of the SDN, which successfully stabilises the plant. Therefore the presented realistic noise model allows us to model impedance control as a result of stochastic OFC. This is an important finding since all previously presented (stochastic) OFC models that used simplistic SDN failed to model such important properties.

Figure 5 shows the discussed effects by overlaying the optimal muscle signal sequence with the induced noise $\mathbf{F}(\mathbf{u})$. We can see that in the naive noise case, any co-contraction would just use more energy while keeping the same noise level, while in our proposed noise scheme, the OFC solution can profit from co-contraction.

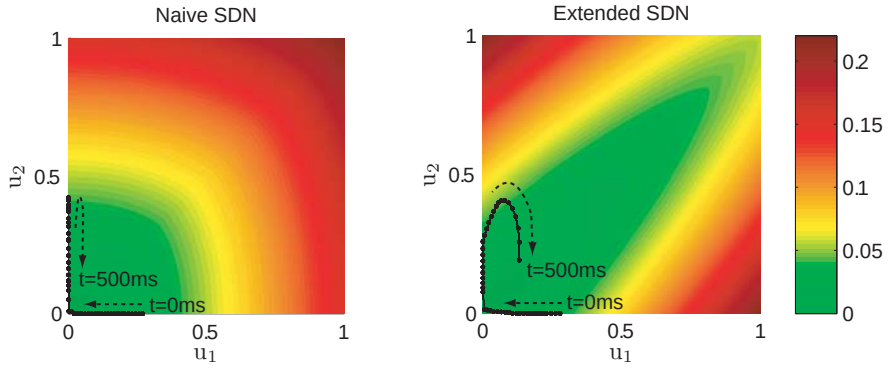


Figure 5: The shaded/coloured regions represent two different control-dependent noise functions $\mathbf{F}(\mathbf{u})$ used in ILQG. Left: $\mathbf{F}(\mathbf{u})$ is a naive SDN. Right: Shaped noise which favours co-contraction. The black lines show the optimal muscle activation sequence found by ILQG, where each dot represents a discrete time step starting at $t = 0$ms and ending at $t = 500$ms. The dashed arrows indicate the time course of the muscle signals.

Next we discuss an internal model representation of the dynamics that allows the system to learn the uncertainty of our plant and can adapt to changes in the plant dynamics.

5 An Internal Model for Uncertainty and Adaptation

It is known that, in order to successfully learn to control a complex system, humans not only need to learn the dynamics of the systems but also about its noise characteristics (Chhabra and Jacobs, 2006). For example an ice hockey player is able to learn that large muscle commands will lead to puck trajectories with large variances over multiple trials, while low muscle commands will lead to trajectories with low variance. The skilled player then can use this additional noise information to create appropriate control policies, for example to make a fast and accurate goal shot.

To learn an approximation $\tilde{\mathbf{f}}$ of the real plant forward dynamics $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u})$ we require a supervised learning method that is capable of non-linear regression. Furthermore we need an efficient incremental method that allows online learning (for adaptation) without suffering from negative interference (Schaal, 2002) and without having to store all previous training data. For such learning problems local methods are particularly well suited. We use *Locally Weighted Projection Regression (LWPR)*, which has been shown to exhibit these beneficial properties and perform very well on motion data (Vijayakumar et al., 2005). Within this local learning paradigm we get access to the stochastic properties of the arm in form of heteroscedastic prediction variances (see Appendix B). As stochastic properties we refer to the kinematic variability of the system described in eq. (11) in Section 2. This induced variability in the training data encodes for uncertainty in the dynamics: if a certain muscle action induces large kinematic

variability over trials this will reduce the variance in those regions. Conversely regions in the state-action space that have little variation will be more trustworthy. Therefore we see that the noise provides additional information about the dynamics of the arm, i.e., the fact that co-contraction makes the limbs more stable and reduces the variability. The noise is predictable because it has been estimated from data produced by the limbs directly, and the motor system might use this information to make the most accurate movement possible (Todorov and Jordan, 2002; Scott, 2004).

We want to continue our investigations with the assumption that our learning system has been pre-trained thoroughly with data from all relevant regions and within the joint limits and muscle activation range of the arm and therefore has acquired an accurate internal model $\tilde{\mathbf{f}}$ of the arm dynamics and its noise properties. Consequently a stochastic OFC problem can be formulated that "guides" the optimal solution towards a *maximum prediction certainty*, while still minimising the energy consumption and end point reaching error. Within OFC we model our dynamics as

$$d\mathbf{x} = \tilde{\mathbf{f}}(\mathbf{x}, \mathbf{u})dt + \mathbf{\Phi}(\mathbf{x}, \mathbf{u})d\boldsymbol{\omega}. \tag{15}$$

This model is analogous to the one presented in Section 4, but the analytic dynamics has been replaced with the learned dynamics $\tilde{\mathbf{f}}$ and the noise model is now represented by $\mathbf{\Phi}(\mathbf{x}, \mathbf{u}) = \sigma^{\mathbf{2}}_{\mathbf{pred}}(\mathbf{x}, \mathbf{u})$. The prediction uncertainty like the dynamics depend on the arm states $\mathbf{x}$ and control actions $\mathbf{u}$. A data driven noise term may also be beneficial when designing anthropomorphic robotic systems, since the modelling of underlying noise processes, originating for example from imperfect hardware, may exhibit complex dependencies on state and actions. Figure 6 shows the proposed stochastic OFC-LD scheme based on a learned internal dynamics, which is an extension of our previous work which only dealt with the deterministic case (Mitrovic et al., 2008a). Notably the internal dynamics model is continuously being updated during reaching with actual data from the arm, allowing the model to account for systematic perturbations, for example due to external force fields (FF).

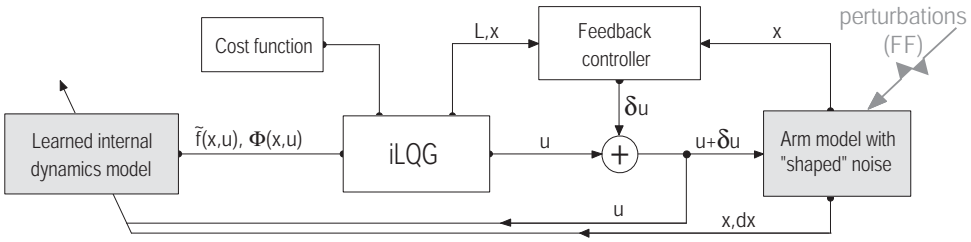


Figure 6: The OFC-LD framework. ILQG produces the optimal control and state sequence, while the optimal feedback matrix L is used to correct deviations from the optimal trajectory optimally.

6 Results

In this Section we show that the proposed OFC-LD framework exhibits viable impedance control, the results of which can be linked to well known patterns of impedance control in human arm reaching. First we will discuss two experiments in stationary dynamics, i.e., the dynamics of the arm and its environment are not changing. The third experiment will model the non-stationary case, where the plant is perturbed by an external force field and the system adapts to the changed dynamics over multiple reaching trials. Before starting the reaching experiments we learned an accurate forward dynamics model $\tilde{\mathbf{f}}$ with data of our arm (for details see Appendix C).

6.1 Experiment 1: Impedance Control for Higher Accuracy Demands

We analyse impedance control in cases where the accuracy demands for reaching of a target are changed while the time for reaching remains constant. In several single and multi-joint experiments an inverse relationship between target size and co-contraction has been observed in humans (Gribble et al., 2003; Osu et al., 2004). As target size is reduced, co-contraction and joint impedance increases and trajectory variability decreases, given that the reaching time remains approximately constant in all conditions. Under these circumstances the CNS takes the energetically more expensive strategy to facilitate arm movement accuracy using higher joint impedance.

To model different accuracy demands in ILQG, we modulate the final cost parameter w_p and w_v in the cost function, which weights the importance of the positional endpoint accuracy and velocity compared to the energy consumption. Like this we create five different accuracy conditions: (A) $w_p = 0.5$, $w_v = 0.25$; (B) $w_p = 1$, $w_v = 0.5$; (C) $w_p = 10$, $w_v = 5$; (D) $w_p = 100$, $w_v = 50$; (E) $w_p = 500$, $w_v = 250$; The energy weight for each condition is $w_e = 1$. Next we used ILQG-LD to simulate optimal reaching starting at $\mathbf{q}_0 = \frac{\pi}{3}$ towards the target $\mathbf{q}_{target} = \frac{\pi}{2}$. Movement time was $T = 500ms$ with a sampling rate of $10ms$ ($dt = 0.01$). For each condition we performed 20 reaching trials.

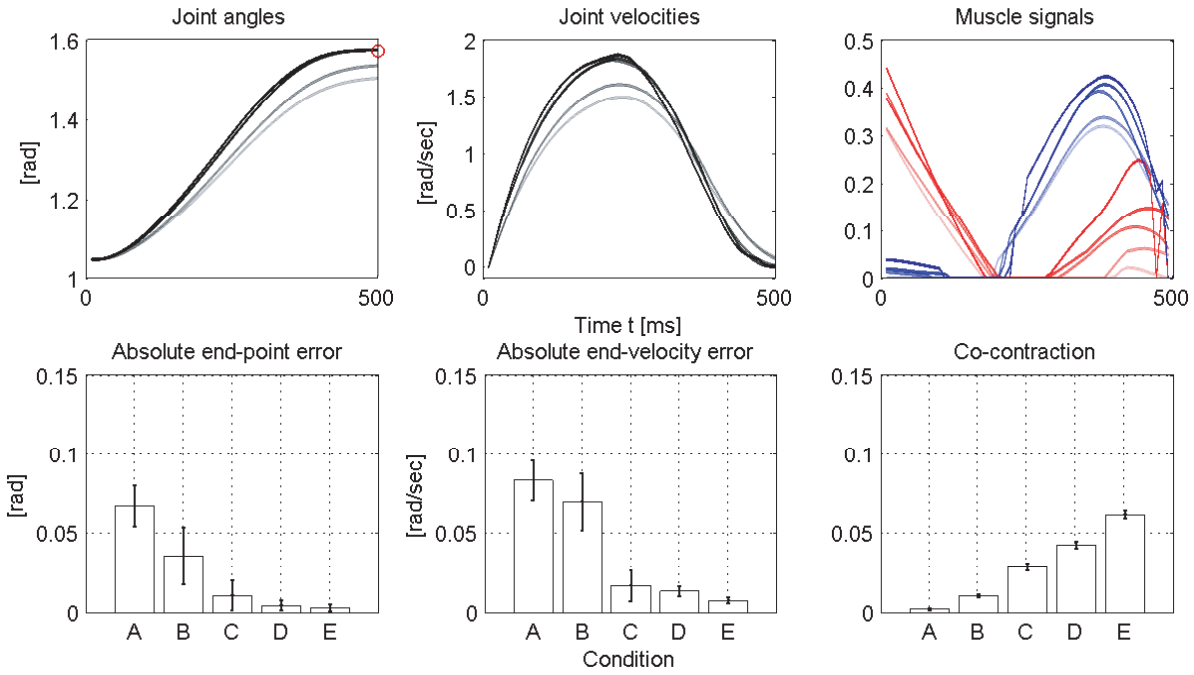


Figure 7: Experimental results from stochastic ILQG-LD for different accuracy demands. The first row of plots shows the averaged joint angles (left), the averaged joint velocities (middle) and the averaged muscle signals (right) over 20 trials for the five conditions A,B,C,D, and E. The darkness of the lines indicates the level of accuracy; the brightest line indicates condition A and the darkest condition E. The bar plots in the second row quantify the reaching performance over 20 trials for each condition. Left: The absolute end-point error and the end-point variability in the trajectories decreases as accuracy demands are increased; Middle: End-point stability also decreases; Right: The averaged co-contraction integrated during 500ms increases with higher accuracy demands, leading to the reciprocal relationship between accuracy and impedance control as observed in humans.

Figure 7 shows the results from the experiment in the five conditions. In condition (A) very low muscle signals are required to satisfy the low accuracy demands, while in the condition

(E), with stringent accuracy demands outweighing the energy term in the cost function, much higher signals are computed. In summary one can observe that if the accuracy demands are increased the muscle signal levels become larger and consequently induce higher co-contraction, which matches well-known neurophysiological results.

6.2 Experiment 2: Impedance Control for Higher Velocities

Here we test our algorithm in conditions where the arm peak velocities are modulated. In humans it has been observed that co-activation increases with maximum joint velocity and it was hypothesised that the nervous system uses a simple strategy to adjust co-contraction and limb impedance in association with movements speed (Suzuki et al., 2001). The causalities here are that faster motion requires higher muscle activity which in turn introduces more noise into the system, the negative effects of which can be limited with higher joint impedance.

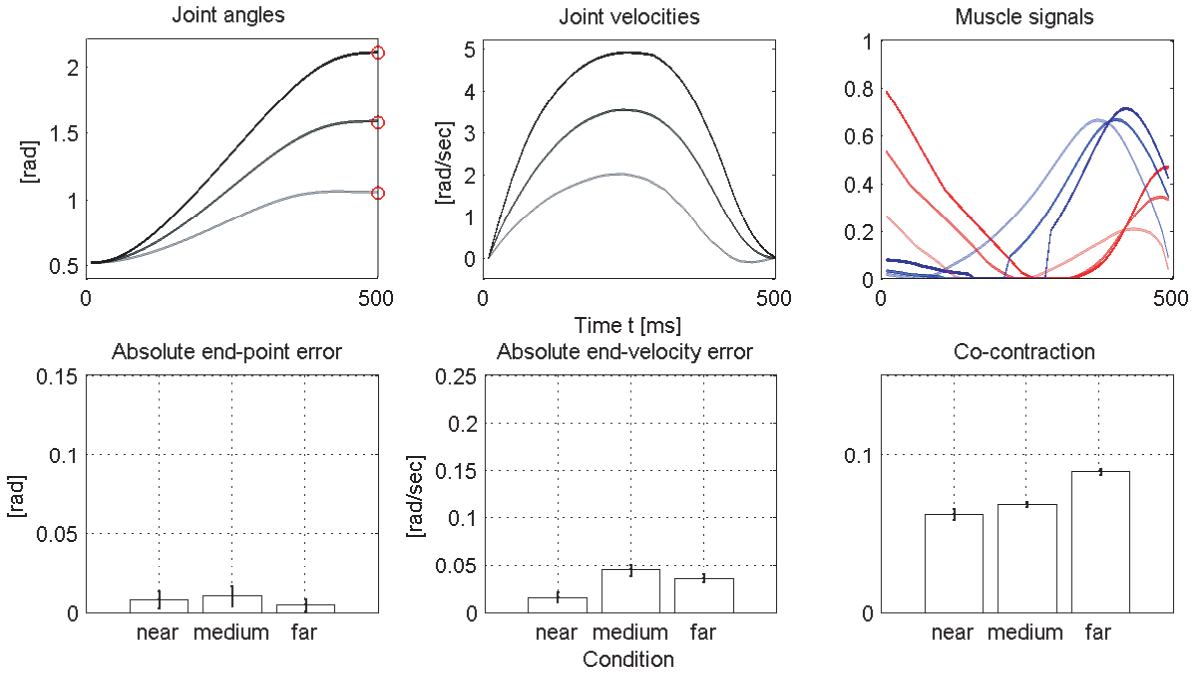


Figure 8: Experimental results from stochastic ILQG-LD for different peak joint velocities. The first row of plots shows the averaged joint angles (left), the averaged joint velocities (middle) and the averaged muscle signals (right) over 20 trials for reaches towards the three target conditions "near", "medium" and "far". The darkest line indicate "far" the brightest indicate the "near" condition. The bar plots in the second row quantify the reaching performance averaged over 20 trials for each condition. The end-point errors (left) and and end-velocity errors (middle) show good performance but no significant differences between the conditions, while co-contraction during the motion as expected increases with higher velocities, due to the higher levels of muscle signals.

As in the previous experiment the reaching time is held constant at $T = 500ms$. We set the new start position to $\mathbf{q}_0 = \frac{\pi}{6}$ and define three reaching targets with increasing distances: $\mathbf{q}_{near} = \frac{\pi}{3}$; $\mathbf{q}_{medium} = \frac{\pi}{2}$; $\mathbf{q}_{far} = \frac{2\pi}{3}$. The cost function parameters are $w_p = 100$, $w_v = 50$, and $w_e = 1$. We again performed 20 trials per condition using ILQG-LD. The results in Figure 8 show that the co-contraction increases for targets that are further away and that have a higher peak velocity. The reaching performance remains solid for all targets, while there are minimal differences in end-point and end-velocity errors between conditions. This can be attributed to

the fact that we reach for different targets, which may be harder or easier to realise for ILQG with the given cost function parameters and reaching time T.

The stationary experiments 1 and 2 exemplified how the proposed stochastic OFC-LD model can explain impedance control that matches well-known psychophysical observations. In both experiments the observed increase in co-contraction can be contributed to the generally higher levels of muscle signals required in the different conditions, which in the OFC-LD framework leads to an increase of co-contraction in order to remain in "more certain" areas of our dynamics model $\tilde{\mathbf{f}}(\mathbf{x}, \mathbf{u})$ and $\mathbf{\Phi}(\mathbf{x}, \mathbf{u})$ respectively. Generally M-shaped co-contraction patterns are produced, which in our particular examples were biased towards the end of the motion. The shape of co-contraction strongly depends on the performed reaching task (i.e. the start and the end positions). Notably such M-shaped stiffness patterns have been reported in humans (e.g., (Gomi and Kawato, 1996)) linking the magnitude of co-activation to the level of reciprocal muscle activation. This supports our OFC-LD methodology as a viable model of how the CNS could cope with SDN.

6.3 Experiment 3: Impedance Control During Adaptation towards External Force Fields

In recent years a large body of experimental work has investigated the motor learning processes in tasks with changing dynamics conditions (e.g., (Burdet et al., 2001; Milner and Franklin, 2005; Franklin et al., 2008)) and it has been shown that subjects generously make use of impedance control to counteract destabilising external forces. In the early stage of dynamics learning humans tend to increase co-contraction and reciprocal muscle activation. As learning progresses in consecutive reaching trials, a reduction in co-contraction with a parallel reduction of the reaching errors made can be observed. While increasing the joint impedance of the arm, subjects were shown to apply lateral forces to counteract the perturbing effect of the force fields. Hence it was hypothesised that the CNS uses an internal model to learn the changes in dynamics and in parallel assists the formation of the dynamics by increasing stability. Consequently impedance control during adaptation tasks can be linked to the uncertainty of the model predictions, as proposed in our model. In the static case the uncertainties are introduced by the neuromotor noise model. Here additional uncertainties are being introduced by a sudden change of the dynamics.

We carried out adaptive reaching experiments with a constant force acting as external perturbation[3]. Within all reaching trials, the ILQG parameters were set to: $T = 500ms$, $w_p = 100$, $w_v = 50$, $w_e = 1$, $\mathbf{q}_0 = \frac{\pi}{2}$, and $\mathbf{q}_{target} = \frac{\pi}{3}$. The varying arm dynamics are modelled using a constant force field $FF = (10, 0, 0)^T$ acting in positive x-direction on the end-effector. As explained in Figure 6, ILQG-LD updates for changes in the dynamics, which leads to characteristic kinematic adaptation patterns observed in humans (Mitrovic et al., 2008a). In the FF catch trial[4], the arm gets strongly deflected when trying to reach for the target because the learned dynamics model $\tilde{\mathbf{f}}(\mathbf{x}, \mathbf{u})$ cannot yet account for the "spurious" forces of the FF. However, using the resultant deflected trajectory as training data and updating the dynamics model online, brings the manipulator nearer to the target with each new trial. Our adaptation experiment starts with 5 trials in the *Null Field (NF)* condition, followed by 20 reaching trials in the FF condition. For each trial we monitored the muscle activations, the co-contraction and the accuracy in the positions and velocities. Because the simulated system is stochastic and never produces exactly the same results (though very similar ones) we repeated the adaptation experiment 20 times under the same conditions and averaged all results. Figure 9 aggregates these results. We see in the kinematic domain (left and middle plots) that the adapted optimal

[3]Many neurophysiological studies deal with multi-joint reaching with rather complex force fields, such as velocity dependent curl fields. We believe that for a basic conceptual understanding the system should be kept as simple and tractable as possible.

[4]The first reach in the new FF condition, which evaluates the effect of the force field before any learning.

solution differs from the NF condition, suggesting that a reoptimisation takes place. After the force field has been learned, the activations for the extensor muscle u_2 are lower and those for the flexor muscle u_1 are higher, meaning that the optimal controller makes use of the supportive force field in positive x-direction. Indeed these results are in line with recent findings in human motor learning, where (Izawa et al., 2008) presented results that suggest that such motor adaptation is not just a process of *perturbation cancellation* but rather a reoptimisation w.r.t. motor cost and the novel dynamics.

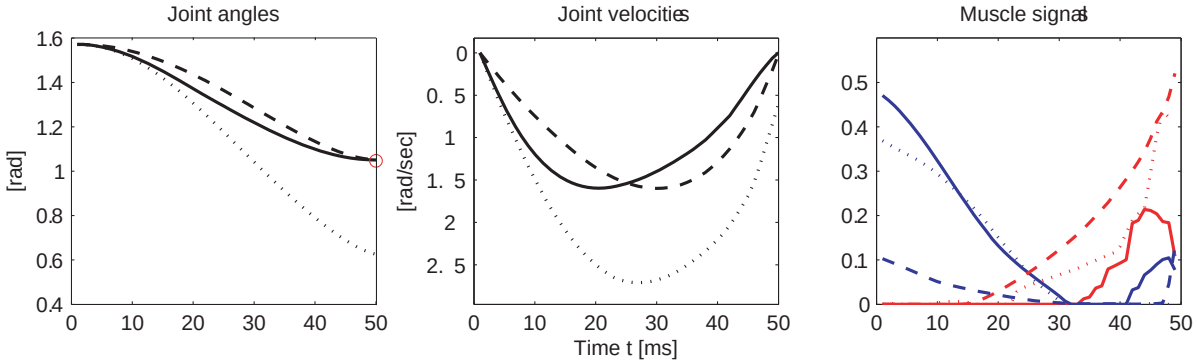


Figure 9: Experimental results from the stochastic ILQG-LD during adaptation. The row of plots shows as before the produced joint angles (left), joint velocities (middle) and muscle signals (right). The solid line represents the reaching with ILQG-LD in the *NF condition*, the dotted shows the *FF catch trial* and the dashed line the trajectories *after adaptation*.

To analyse the adaptation process in more detail, Figure 10 presents the integrated muscle signals and co-contraction, the produced absolute end-point and end-velocity errors and the prediction uncertainty (i.e., LWPR confidence bounds) during each of the performed 25 reaching trials. The confidence bounds were computed after each trial with the updated dynamics along the current trajectory. The first five trials in the NF show approximately constant muscle parameters along with good reaching performance and generally low confidence bounds. Even in the NF condition the learning further decreases the already low confidence bounds. In trial 6, the FF catch trial, the reaching performance decrease drastically due to the novel dynamics. This also increase the confidence bounds since the input distribution along the current trajectory has changed and "blown up" the confidence bounds in that region. Consequently ILQG-LD now faces wider confidence bound along that new trajectory. These can be reduced by increasing co-contracting and therefore enter lower noise regions, which allow the algorithm to keep the confidence bounds lower and still produce enough joint torque. For the next four trials (i.e., trials 7 to 10) the co-activation level stays elevated, while the internal model gets updated, which is indicated by the change in reciprocal activations and improved performance between the trials. After the 11^{th} trial the impedance has reduced to roughly the normal NF level and the confidence bound are fairly low (< 1) and keep decreasing, which shows the expected connection between impedance and prediction uncertainty.

A further indication for the viability of our impedance control is a direct comparison to the deterministic case. We repeated exactly the same adaptation experiment using the deterministic ILQG-LD, meaning the algorithm ignored the stochastic information available through the confidence bounds (Figure 11). For the deterministic case one can observe that as in the static experiments (Section 4) virtually no co-contraction during adaptation is produced. This leads generally to larger errors in the early learning phase phase (trial 6 to 10), especially in the joint velocities. In contrast, for the stochastic algorithm the increased joint impedance stabilises the arm better towards the effects of the FF and therefore produces smaller errors.

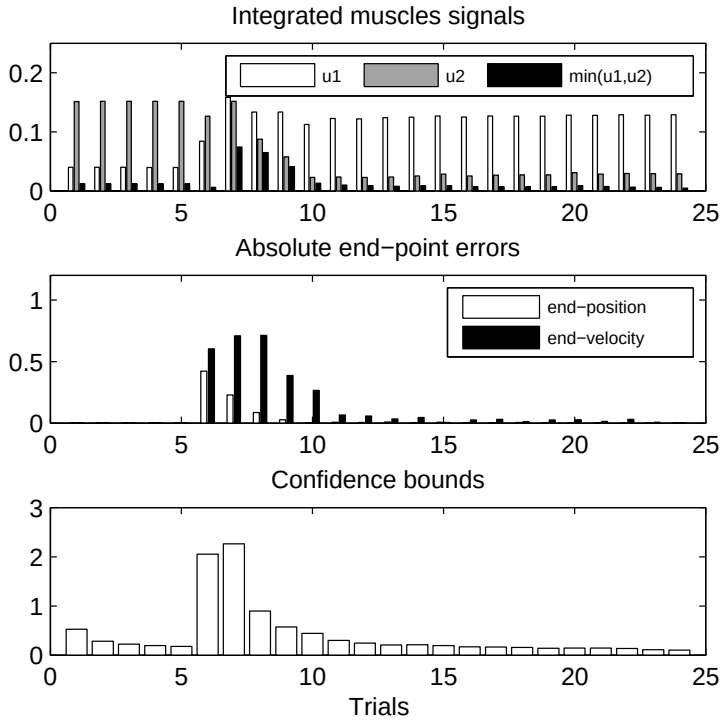


Figure 10: Accumulated statistics during 25 adaptation trials using stochastic ILQG-LD. Trials 1 to 5 are performed in the NF condition. Top: Muscle parameters and co-contraction integrated during 500ms reaches. Middle: Absolute joint errors and velocity errors at final time $T = 500ms$. Bottom: Integrated confidence bounds along the current optimal trajectory after this has been learned.

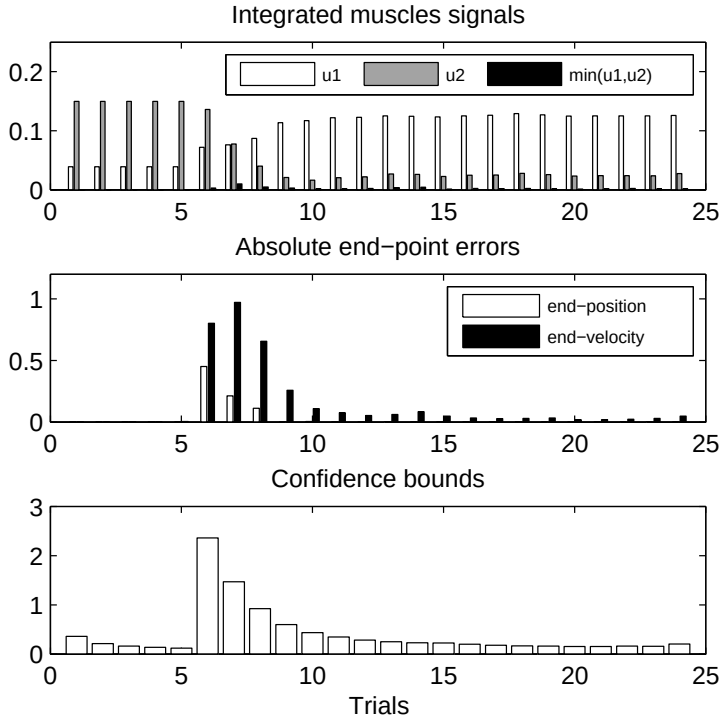


Figure 11: Accumulated statistics during 25 adaptation trials using *deterministic* ILQG-LD.

7 Discussion

In summary this paper presented a model for joint impedance control, which is a key strategy of the CNS to produce limb control that is stable towards internal and external fluctuations. Our model is based on the fundamental assumption that the CNS, besides optimising for energy and accuracy, maximises the certainty from its internal dynamics model predictions. We hypothesized that, in conjunction with an appropriate antagonistic arm and signal-dependent noise model, impedance control emerges as a result from an optimisation process. We formalised our model within the theory of stochastic OFC, in which an actor's goal is expressed as a solution to an optimisation process that minimizes energy consumption and end-point error. Such optimal control problems can be solved efficiently via iterative local methods like ILQG. Unlike previous OFC models we made the assumption that the actor utilises a learned dynamics model from data that are produced by the system directly. The plant data itself was generated from a stochastic arm model that we developed in such a way that its noise-impedance characteristics resemble the ones of humans. Like this we introduced a signal dependent kinematic variability into the system, the magnitude of which decreases with higher co-activation levels. The learner interprets this kinematic variability, here also termed noise, as prediction uncertainty which is given algorithmically in form of heteroscedastic confidence bounds within LWPR. With these ingredients we formulated a stochastic OFC algorithm that uses the learned dynamics and the contained uncertainty information; we named this algorithm ILQG-LD. This general model for impedance control of antagonistic limb systems is based on the quality of the learned internal model and therefore leads to the intuitive requirement that impedance will be increased in cases where the actor is uncertain about the model predictions. We evaluated our model in several simulation experiments, both in stationary and adaptation tasks. The results showed that our general model can explain numerous experimental findings from the neurophysiology literature.

Is OFC-LD a valid model for how the CNS manages impedance control?

Besides the experimental evidence, OFC-LD seems a plausible approach for modelling how the CNS realises impedance control. The formulation within optimal control theory is well motivated, since it explains many biologically relevant factors like the integrated redundancy resolution and trajectory generation, the motion patterns and observed motor synergies. Indeed in recent years optimality has become the predominant theoretical framework to the modelling of volitional motor control. Furthermore, our model utilises the concepts of learned internal dynamics models, which without doubt plays an important role in the formation of sensorimotor control strategies and which in practice allowed us to model adaptation and re-optimisation behaviour as shown in experiment 3. Notably our learning framework uses exclusively data for learning that are available to the biological motor system through visual and proprioceptive feedback, therefore delivering a sound analogy to a human learner. A fundamental aspect this work addressed, is that the signal dependent noise information can provide additional information about the system dynamics, which may be an indication of the possible constructive role of noise in the neuromotor system (Faisal et al., 2008). Even though there is limited knowledge about the neural substrate behind the observed optimality principles in motor control (Shadmehr and Krakauer, 2008), by using the learned internal model paradigm, we believe to have pushed the OFC theory nearer towards a more plausible neural representation. Our model has shown, for the first time, that impedance control can be interpreted as an optimisation process, therefore providing further support for the important role of optimality principles in human motor control.

What are the drawbacks of OFC-LD?

From a computational perspective, the ILQG algorithm currently seems to be the most suitable algorithm available to find OFC laws, and in practice it already showed to scale well to more complex systems (Mitrovic et al., 2008b; Mitrovic et al., 2008a). A limiting factor though is the dynamics learning using local methods, which to a certain extent suffer from the curse of dimensionality, in that the learner has to produce a vast amount of training data to cover the whole state-action space. Another drawback of stochastic OFC-LD is that so far we only modelled process noise and ignored observation noise entirely. This is a simplification to real natural systems, in which large noise in the observations is present, both from vision and proprioceptive sensors (Faisal et al., 2008).

Future work

For the future more detailed biomechanical models could lead to a better understanding of the noise-impedance characteristics of human limbs, for isometric and isotonic contractions. Gathering such data in practice is a rather involved and difficult task and the future development in that field will need to be observed closely. A useful extension to the noise model in OFC-LD would contain the introduction of state dependencies in $\mathbf{F}(\mathbf{x}, \mathbf{u})$, which for example could model the known important muscle length dependencies. State-dependent noise would further give ways to model effects of the apparent sensorimotor delays, which were not addressed in this paper. A human learner receives data caused by delayed feedback, and for movements with large velocities the negative effects of feedback delays are more apparent than for slow motion. Such dependencies could in theory be modelled as state-dependent noise, which in this example would be reflected as velocity-dependent variability in the training data.

Other future work will need to address the shortcomings of our current arm model. We used a single-joint arm for an easier understanding and better computational tractability. In order to improve the comparability of our results to current neurophysiological findings a more accurate biomechanical arm model should be implemented that follows multi-joint kinematics, a muscle model with multiple (biarticulate) muscle groups, and more realistic tension properties using muscle activation dynamics. We believe that the presented results, from the stationary and adaptive experiments, will scale to higher DoF systems, since impedance control originates from the antagonistic muscle structure in the joint-space domain. Obviously the used muscle and kinematics model will determine to a large extent the produced joint impedance characteristics and its transformation into the Cartesian end-effector domain (i.e., task space). It remains to be seen whether the stochastic OFC-LD framework has the capability to explain other important multi-joint impedance phenomena such as the end-effector stiffness (Burdet et al., 2001) that is selectively tuned towards the directions of instability.

Appendix A: The ILQG algorithm

The ILQG algortihm starts with a time-discretised initial guess of an optimal control sequence and then iteratively improves it w.r.t. the performance criteria in v (eq. 13). From the initial control sequence $\bar{\mathbf{u}}^i$ at the i-iteration, the corresponding state sequence $\bar{\mathbf{x}}^i$ is retrieved using the deterministic forward dynamics $\mathbf{f}$ with a standard Euler integration $\bar{\mathbf{x}}^i_{k+1} = \bar{\mathbf{x}}^i_k + \Delta t\, \mathbf{f}(\bar{\mathbf{x}}^i_k, \bar{\mathbf{u}}^i_k)$. In a next step the discretised dynamics (eq. 12) are linearly approximated and the cost function (eq. 14) is quadratically approximated around $\bar{\mathbf{x}}^i_k$ and $\bar{\mathbf{u}}^i_k$. The approximations are formulated as deviations of the current optimal trajectory $\delta\mathbf{x}^i_k = \mathbf{x}^i_k - \bar{\mathbf{x}}^i_k$ and $\delta\mathbf{u}^i_k = \mathbf{u}^i_k - \bar{\mathbf{u}}^i_k$ and therefore form a "local" LQG problem. This linear quadratic problem can be solved efficiently via a

modified Ricatti-like set of equations. The optimisation supports constraints for the control variable $\mathbf{u}$, such as lower and upper bounds. After the optimal control signal correction $\delta\bar{\mathbf{u}}^i$ has been obtained, it can be used to improve the current optimal control sequence for the next iteration using $\bar{\mathbf{u}}_k^{i+1} = \bar{\mathbf{u}}_k^i + \delta\bar{\mathbf{u}}^i$. At last $\bar{\mathbf{u}}_k^{i+1}$ is applied to the system dynamics (eq. 12) and the new total cost along the along the trajectory is computed. The algorithm stops once the cost v cannot be significantly decreased anymore. After convergence, ILQG returns an optimal control sequence $\bar{\mathbf{u}}$ and a corresponding optimal state sequence $\bar{\mathbf{x}}$ (i.e., trajectory). Along with the optimal open loop parameters $\bar{\mathbf{x}}$ and $\bar{\mathbf{u}}$, ILQG produces a feedback matrix $\mathbf{L}$ which may serve as optimal feedback gains for correcting local deviations from the optimal trajectory on the plant.

Appendix B: Outline of the LWPR algorithm

In LWPR, the regression function is constructed by blending local linear models, each of which is endowed with a locality kernel that defines the area of its validity (also termed its receptive field). During training, the parameters of the local models (locality and fit) are updated using incremental Partial Least Squares, and models can be pruned or added on an as-need basis, for example, when training data is generated in previously unexplored regions. Usually the receptive fields of LWPR are modelled by Gaussian kernels, so their activation or response to a query vector $\mathbf{z}$ (combined inputs $\mathbf{x}$ and $\mathbf{u}$ of the forward dynamics $\tilde{\mathbf{f}}$) is given by

$$w_k(\mathbf{z}) = \exp\left(-\frac{1}{2}(\mathbf{z}-\mathbf{c}_k)^T\mathbf{D}_k(\mathbf{z}-\mathbf{c}_k)\right), \tag{16}$$

where $\mathbf{c}_k$ is the centre of the k^{th} linear model and $\mathbf{D}_k$ is its distance metric. Treating each output dimension separately for notational convenience, the regression function can be written as

$$\tilde{f}(\mathbf{z}) = \frac{1}{W}\sum_{k=1}^{K} w_k(\mathbf{z})\psi_k(\mathbf{z}), \quad W = \sum_{k=1}^{K} w_k(\mathbf{z}), \tag{17}$$

$$\psi_k(\mathbf{z}) = b_k^0 + \mathbf{b}_k^T(\mathbf{z}-\mathbf{c}_k), \tag{18}$$

where b_k^0 and $\mathbf{b}_k$ denote the offset and slope of the k-th model, respectively.

LWPR learning has the desirable property that it can be carried out online, and moreover, the learned model can be adapted to changes in the dynamics in real-time. A forgetting factor λ (Vijayakumar et al., 2005), which balances the trade-off between preserving what has been learned and quickly adapting to the non-stationarity, can be tuned to the expected rate of external changes.

The statistical parameters of LWPR regression models provide access to the confidence intervals, here termed *confidence bounds*, of new prediction inputs (Vijayakumar et al., 2005). In LWPR the predictive variances are assumed to evolve as an additive combination of the variances within a local model and the variances independent of the local model. The predictive variance estimates $\sigma_{pred,k}^2$ for the k-th local model can be computed in analogy with ordinary linear regression. Similarly one can formulate the global variances σ^2 across models. In analogy to eq. (17) LWPR then combines both variances additively to form the confidence bounds given by

$$\sigma_{pred}^2 = \frac{1}{W^2}\left(\sum_{k=1}^{K} w_k(\mathbf{z})\sigma^2 + \sum_{k=1}^{K} w_k(\mathbf{z})\sigma_{pred,k}^2\right). \tag{19}$$

The local nature of LWPR leads to the intuitive requirement that only receptive fields that actively contribute to the prediction (e.g., large linear regions) are involved in the actual confidence bounds calculation. Large confidence bound values typically evolve if the training data contains much noise and other sources of variability such as changing output distributions.

Further regions with sparse or no training data, i.e. unexplored regions, show large confidence bounds compared to densely trained regions. Figure 12 depicts the learning concepts of LWPR graphically on a learned model with one input and one output dimension. The noisy training data was drawn from an example function that becomes more linear and more noisy for larger $z-$values. Furthermore in the range $z = [5..6]$ no data was sampled for training to show the effects of sparse data on LWPR learning.

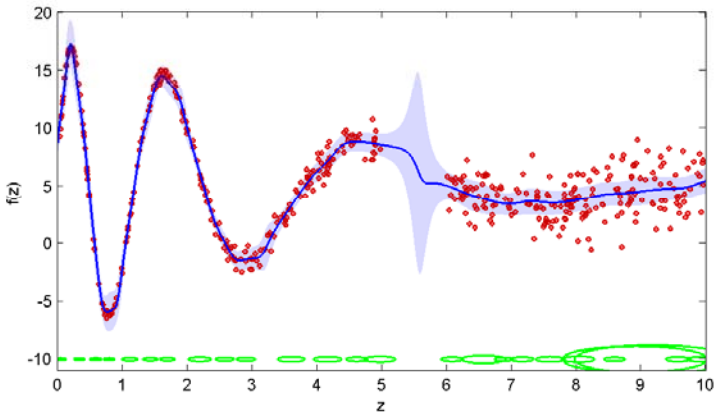


Figure 12: Typical regression function (blue continuous line) using LWPR. The dots indicate a representative training data set. The receptive fields are visualised as ellipses drawn at the bottom of the plot. The shaded region represents the confidence bounds around the prediction function. The confidence bounds grow between $\mathbf{z} = [5..6]$ (no training data) and generally towards larger $\mathbf{z}$ values (noise grows with larger values).

Appendix C: Details on Learning of the Internal Model

We coarsely pre-trained an LWPR dynamics model with a data set S collected from the arm model *without* using the extended noise model. The data was densely and randomly sampled from the arm's operation range with $\mathbf{q} = [\frac{2}{9}\pi, \frac{7}{9}\pi], \dot{\mathbf{q}} = [-2\pi, 2\pi]$, and $\mathbf{u} = [0, 1]$. The collected data set ($2.5 \cdot 10^6$ datapoints) was split into a 70% training set and a 30% test set. We stopped learning once the model prediction of $\tilde{\mathbf{f}}(\mathbf{x}, \mathbf{u})$ could accurately replace the analytic model $\mathbf{f}(\mathbf{x}, \mathbf{u})$, which was checked using the *normalised mean squared error (nMSE)* of $5 \cdot 10^{-4}$ on the test data. After having acquired the noise free dynamics accurately we collected a second data set S^{noise} in analogy to S but this time the data was drawn from the arm model *including* the extended noise model. We then used S^{noise} to continue learning on on our existing dynamics model $\tilde{\mathbf{f}}(\mathbf{x}, \mathbf{u})$. The second learning round has primarily the effect of shaping the confidence bounds according to the noise in the data and the learning is stopped once the confidence bounds stop changing. One correctly could argue that such a two step learning approach is biologically not feasible because a human learning system for example never gets noise-free data. The justification of our approach is of a practical nature and simplifies the rather involved initial parameter tuning of LWPR and allows us to monitor the global learning success (via the nMSE) more reliably over the large data space. Fundamentally though, our learning method does not conflict with any stochastic OFC-LD principles that we proposed.

References

Burdet, E., Osu, R., Franklin, D. W., Milner, T. E., and Kawato, M. (2001). The central nervous system stabilizes unstable dynamics by learning optimal impedance. *Nature*, 414:446–449.

Burdet, E., Tee, K. P., Mareels, I., Milner, T. E., Chew, C. M., Franklin, D. W., Osu, R., and Kawato, M. (2006). Stability and motor adaptation in human arm movements. *Biological Cybernetics*, 94:20–32.

Chhabra, M. and Jacobs, R. A. (2006). Near-optimal human adaptive control across different noise environments. *Journal of Neuroscience*, 26(42):10883–10887.

Davidson, P. R. and Wolpert, D. M. (2005). Widespread access to predictive models in the motor system: a short review. *Journal of Neural Engineering*, 2:313–319.

Faisal, A. A., Selen, L. P. J., and Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, 9:292–303.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47:381–391.

Franklin, D. W., Burdet, E., Tee, K. P., Osu, R., Chew, C.-M., Milner, T. E., and Kawato, M. (2008). Cns learns stable, accurate, and efficient movements using a simple algorithm. *Journal of Neuroscience*, 28(44):11165–11173.

Gomi, H. and Kawato, M. (1996). Equilibrium-point control hypothesis examined by measured arm stiffness during multijoint movement. *Science*, 272(5258):117–120.

Gribble, P. L., Mullin, L. I., Cothros, N., and Mattar, A. (2003). Role of cocontraction in arm movement accuracy. *Journal of Neurophysiology*, 89(5):2396–2405.

Hamilton, A. and Wolpert, D. M. (2002). Controlling the statistics of action: Obstacle avoidance. *Journal of Neurophysiology*, 87:2434–2440.

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

Haruno, M. and Wolpert, D. M. (2005). Optimal control of redundant muscles in step-tracking wrist movements. *Journal of Neurophysiology*, 94:4244–4255.

Hogan, N. (1984). Adaptive control of mechanical impedance by coactivation of antagonist muscles. *IEEE Transactions on Automatic Control*, 29(8):681–690.

Izawa, J., Rane, T., Donchin, O., and Shadmehr, R. (2008). Motor adaptation as a process of reoptimization. *Journal of Neuroscience*, 28(11):2883–2891.

Jacobson, D. H. and Mayne, D. Q. (1970). *Differential Dynamic Programming*. Elsevier, New York.

Jones, K. E., Hamilton, A. F., and Wolpert, D. M. (2002). Sources of signal-dependent noise during isometric force production. *Journal of Neurophysiology*, 88(3):1533–1544.

Katayama, M. and Kawato, M. (1993). Virtual trajectory and stiffness ellipse during multijoint arm movement predicted by neural inverse model. *Biological Cybernetics*, 69:353–362.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, 9(6):718–727.

Li, W. (2006). *Optimal Control for Biological Movement Systems*. PhD dissertation, University of California, San Diego.

Li, W. and Todorov, E. (2004). Iterative linear-quadratic regulator design for nonlinear biological movement systems. In *Proc. 1st Int. Conf. Informatics in Control, Automation and Robotics*.

Liu, D. and Todorov, E. (2007). Evidence for the flexible sensorimotor strategies predicted by optimal feedback control. *Journal of Neuroscience*, 27:9354–9368.

Milner, T. E. and Franklin, D. (2005). Impedance control and internal model use during the initial stage of adaptation to novel dynamics in humans. *Journal of Physiology*, 567:651–664.

Mitrovic, D., Klanke, S., and Vijayakumar, S. (2008a). Adaptive optimal control for redundantly actuated arms. In *In proceedings of the 10th International Conference on Simulation of adaptive behaviour (SAB)*, Osaka, Japan.

Mitrovic, D., Klanke, S., and Vijayakumar, S. (2008b). Optimal control with adaptive internal dynamics models. In *In proceedings of the 5th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, Madeira, Portugal.

Osu, R. and Gomi, H. (1999). Multijoint muscle regulation mechanisms examined by measured human arm stiffness and emg signals. *Journal of Neurophysiology*, 81:1458–1468.

Osu, R., Kamimura, N., Iwasaki, H., Nakano, E., Harris, C. M., Wada, Y., and Kawato, M. (2004). Optimal impedance control for task achievement in the presence of signaldependent noise. *Journal of Neurophysiology*, 92(2):1199–1215.

Perreault, E. J., Kirsch, R. F., and Crago, P. E. (2001). Effects of voluntary force generation on the elastic components of endpoint stiffness. *Experimental Brain Research*, 141:312–323.

Schaal, S. (2002). *The handbook of brain theory and neural networks*, chapter Learning Robot Control, pages 983–987. MIT Press, Cambridge, MA.

Scott, S. H. (2004). Optimal feedback control and the neural basis of volitional motor control. *Nature Reviews Neuroscience*, 5:532–546.

Selen, L. P., Beek, P. J., and Dieen, J. H. (2005). Can co-activation reduce kinematic variability? a simulation study. *Biological Cybernetics*, 93:373–381.

Selen, L. P. J. (2007). *Impedance modulation: a means to cope with neuromuscular noise*. Phd thesis, Amsterdam: Vrije Universiteit.

Shadmehr, R. and Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Experimental Brain Research*, 185:359–381.

Stengel, R. F. (1994). *Optimal control and estimation*. Dover Publications, New York.

Suzuki, M., Douglas, M. S., Gribble, P. L., and Ostry, D. J. (2001). Relationship between cocontraction, movement kinematics and phasic muscle activity in single-joint arm movement. *Experimental Brain Research*, 140:171–181.

Tee, K. P., Burdet, E., Chew, C. M., and Milner, T. E. (2004). A model of force and impedance in human arm movements. *Biological Cybernetics*, 90:368–375.

Thoroughman, K. A. and Shadmehr, R. (1999). Electromyographic correlates of learning an internal model of reaching movements. *Journal of Neuroscience,*, 19(19):8573–8588.

Todorov, E. (2005). Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. *Neural Computation*, 17(5):1084–1108.

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

Todorov, E. and Li, W. (2005). A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems. In *Proc. of the American Control Conference*.

Vijayakumar, S., D'Souza, A., and Schaal, S. (2005). Incremental online learning in high dimensions. *Neural Computation*, 17:2602–2634.

Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, 269(5232):1880–1882.