

Inverse Optimality Design for Biological Movement Systems

Weiwei Li* Emanuel Todorov** Dan Liu***

* *Nordson Asymtek, Carlsbad, CA 92010 USA (e-mail: wwli@ieee.org).*

** *University of Washington, Seattle, WA 98195 USA*

*** *Google, Mountain View, CA 94043 USA*

Abstract: This paper proposes an inverse optimality design method for nonlinear deterministic system and nonlinear stochastic system with multiplicative and additive noises. The new method is developed based on the general Hamilton-Jacobi-Bellman(HJB) equation, and it constructs an estimated cost function using a linear function approximator—Gaussian Radial Basis function, which can recover the original performance criterion for which the given control law is optimal. The performance of the algorithm is illustrated on scalar systems and a complex biomechanical control problem involving a stochastic model of the human arm.

Keywords: Inverse dynamic problem, Inverse optimal control, Biological movement.

1. INTRODUCTION

When we are examining the human movement, such as reaching, throwing, orienting, it is usually hypothesized that natural control systems may be implicitly optimizing an objective function. However, it is not known what this performance criterion might be. In order to establish the correspondence between the natural and the artificial systems, we need to ascertain what it is that the natural control system is optimizing. Investigations of general optimal control problems assume that the performance criteria are known. Based on this performance criterion and a given dynamic system, an optimal feedback control law is derived to satisfy the Hamilton-Jacobi-Bellman(HJB) equation. The inverse problem of optimal control, on the other hand, solves for the coefficients of cost function for which a given control law is optimal. Surprisingly, the novel inverse optimal control approach gives us a clue to determine the cost function by fitting the mathematical model to real behavior in a variety of circumstances.

The viewpoint of inverse optimum control problem was originated in the early 1960s by Kalman (1964): "Given a dynamic system and a feedback control law, which is known to yield an asymptotically stable closed-loop system, the inverse problem is to seek the most general performance index for which this control law is optimal." Kalman considered a very precise formulation of this inverse problem for linear time-invariant single-input systems and derived a beautiful conclusion for linear quadratic regulators (Anderson et al. (1990)). For a given control law $u = -kx$ of the deterministic linear system $\dot{x} = Fx + gu$, a necessary and sufficient condition for this control law to be optimal with respect to a performance criterion $\int_{t_0}^{\infty} (u^2 + x^T Q x) dt$ is that the return difference condition is hold for all real frequencies.

The inverse optimal control approach was then introduced into nonlinear systems by Thau (1967), Yokoyama et al. (1972) and Moylan et al. (1973). The main results of

Thau (1967) are restricted to system where the optimum performance criterion is required to have a quadratic term in the state variables. Yokoyama et al. (1972) extent results for a general class of multi-input systems, and found the necessary and sufficient conditions for optimized performance indices. While Moylan et al. (1973) considered another class of inverse problems that the control variables are quadratic in the performance index, but not necessary the state, and established a certain nonlinear return inequality which implies robustness to some input nonlinearities. To our knowledge, this inverse optimality method has not been well-studied until it was revived for the design of robust nonlinear controllers by Freeman et al. (1996) and Sepulchre et al. (1997). Instead of designing a controller for a given cost functional, the inverse optimal control problem in Freeman et al. (1996) and Sepulchre et al. (1997) was concerned with seeking a controller that optimizes some meaningful cost functionals. The main contribution of the inverse optimality approach developed by Freeman et al. (1996) is to construct a control Lyapunov function as an optimal value function and then derive a stabilizing control law which is optimal with respect to this value function. This approach has then been employed to construct optimal feedback control laws for rigid spacecraft by Krstic et al. (1999) and some other stochastic nonlinear systems, see Deng et al. (1997), Deng et al. (1999), Li et al. (1997), Tang et al. (2002) and Dvijotham et al. (2010). One way of approaching optimal control problems of stochastic nonlinear system involves solving the Hamilton-Jacobi-Bellman (HJB) partial differential equation, which turns out to be an infeasible task. The benefit of the inverse optimal control approach is that it avoids the burden of solving HJB equation, whenever the optimal cost-to-go function is known (i.e. inferred from the data), it becomes an explicit formula and results in a controller optimal with respect to a meaningful cost functional.

The aim of this research is to address the inverse problem of optimal control to the biological movement systems,

that is, the problem of extracting the objective function that can explain the observed, optimal movement behavior. Such a goal is possible because of significant experimental advances in our understanding of motor control systems in humans, and because of the increased knowledge in computational models for control learning, see Todorov et al. (2002), Todorov et al. (2005) and Li et al. (2007). It will create a new way not only to analyze the natural movement phenomena, but also to derive much more effective methods in modelling all kinds of purposive(or skilled) behavior.

In this paper, we focus on the study of inverse optimality design for nonlinear system. Given a nonlinear dynamic system and a feedback control law optimized with respect to a discounted cost function (which is quadratic on the control, but not the state), we develop a new algorithm to construct an estimated cost function which can recover the original performance criterion for which this control law is optimal. The new method used the general HJB equation which can deal with nonlinear dynamic cases, and our approximation to cost function uses Gaussian basis functions which do not assume a quadratic form. We shall show that the simulation results clearly demonstrate the effectiveness and advantages of the proposed approach.

The paper is organized as follows. In section 2 the inverse optimality problem is formulated for a class of nonlinear deterministic system and performance indices. An algorithm for the construction of cost functional is also developed in section 2. Section 3 addresses the inverse optimality design for a class of nonlinear stochastic system. Section 4 presents several numerical examples: one is on the inverse optimal control problem of the deterministic/stochastic linear system under the quadratic cost assumption, the other is the application to human movement control. Concluding remarks are drawn in section 5.

2. INVERSE DESIGN FOR DETERMINISTIC SYSTEM

2.1 Problem Formulation

Consider the problem of minimizing the infinite horizon discounted cost function subject to the nonlinear dynamic system

$$\dot{x} = f(x) + G(x)u, \quad (1)$$

where $x \in \mathbf{R}^{n_x}$ and $u \in \mathbf{R}^{n_u}$ are state and control variables, $f(x)$ and $G(x)$ are known nonlinearities. The performance criterion to be minimized is

$$V(x(0)) = \int_0^\infty e^{-t/\tau} \left[q(x(t)) + \frac{1}{2}u(t)^T R u(t) \right] dt, \quad (2)$$

where $q(x)$ is a smooth function of its arguments, $R \in \mathbf{R}^{n_u \times n_u}$ is symmetric and positive definite, and $\tau > 0$ represents the constant discount factor.

The inverse optimal control problem can now be formulated as follows. Given the dynamic system (1) and a feedback control law $u(t)$ as

$$u(t) = \Psi(x(t)) \quad (3)$$

which can result in an asymptotically stabilized closed-loop system

$$\dot{x} = f(x) + G(x)\Psi(x), \quad (4)$$

the objective here is to recover the cost function (including $q(x)$ and R) for which the control law (3) is optimal.

2.2 Cost function Estimation

Let us define the Hamiltonian \mathcal{H} as

$$\begin{aligned} \mathcal{H}(x(t), u(t), V(x(t))) &= q(x(t)) + \frac{1}{2}u(t)^T R u(t) \\ &\quad + V_x^T \left(f(x) + G(x)u(t) \right), \end{aligned}$$

where V_x is a simplified notation of partial derivative of $V(x(t))$ over x . A necessary and sufficient condition for a control law

$$u^*(t) = \Psi(x(t)) = -R^{-1}G(x)^T V_x^*(x(t)) \quad (5)$$

to be optimal is that it satisfies the following optimality condition

$$\frac{1}{\tau} V^*(x(t)) = \min_{u(t)} \left\{ \mathcal{H}(x(t), u(t), V^*(x(t))) \right\}. \quad (6)$$

The above partial differential equation (6) is called the Hamilton-Jacobi-Bellman (HJB) equation, where $V^*(x(t))$, known as the optimal value function (optimal cost-to-go), is the accumulated expected cost if the system is initialized in state x at time t , and the optimal control law $u^*(t)$ is applied until the end of optimal control history; V_x^* is the partial derivative of $V^*(x(t))$ over x . Bellman's optimality principle refers to $V(x(t))$ and $u(t)$ at all possible states $x(t)$, and therefore leads to global methods that compute an optimal control law.

A necessary condition for $u(t)$ to minimize \mathcal{H} is that

$$\frac{\partial \mathcal{H}}{\partial u} = 0, \quad (7)$$

thus the optimal control (5) is obtained by solving the above equation (7).

Substituting (5) into (6), we obtain the HJB equation as

$$\begin{aligned} \frac{1}{\tau} V^*(x(t)) &= q(x(t)) + V_x^{*T} f(x) \\ &\quad - \frac{1}{2} V_x^{*T} G(x) R^{-1} G^T(x) V_x^*. \end{aligned} \quad (8)$$

In order to obtain more explicit results on the value of the optimal performance index, we use the function approximation method to estimate the optimal value function

$$V^*(x(t)) \cong \sum_{i=1}^m w_i \Phi_i(x), \quad (9)$$

where

$$\Phi_i(x) = \exp \left(-\frac{\|x - c_i\|^2}{r_i^2} \right) \quad (10)$$

is the i^{th} Gaussian radial basis function (GRBF), c_i defines its center and r_i effectively determines the width of GRBF; w_i is the associated weight of $\Phi_i(x)$. By defining $\mathbf{w} \in \mathbf{R}^m$ as a vector of linear weights

$$\mathbf{w} = (w_1 \ w_2 \ \cdots \ w_m)^T,$$

and $\Phi(x) \in \mathbf{R}^m$ as a vector of Gaussian radial basis functions

$$\Phi(x) = \left(\Phi_1(x) \ \cdots \ \Phi_m(x) \right)^T,$$

the optimal value function (9) can be written as follows

$$V^*(x(t)) = \mathbf{w}^T \Phi(x). \quad (11)$$

Actually the Gaussian radial basis function $\Phi(\cdot)$ plays an important role in traditional neural networks, and is often used in reinforcement learning to learn value function approximations having local generalization properties. Here the optimal value function (11) is expressed as a linear combination of a set of m basis functions. By taking the derivative over x on both sides of (11), it yields

$$V_x^*(x(t)) = \mathbf{w}^T \Phi_x(x), \quad (12)$$

where $\Phi_x(x)$ defines the derivative of $\Phi(x)$ over x . Substituting (12) into (5), we have

$$u^*(t) = -R^{-1}G(x)^T \Phi_x^T(x) \mathbf{w}. \quad (13)$$

By choosing R as a fixed positive definite matrix and assuming $R^{-1}G(x)^T \Phi_x^T(x)$ is rank efficient, the linear parameters \mathbf{w} of the Gaussian radial basis functions can then be determined by solving the above equation using the least squares regression method.

Finally, substituting (11) and (12) into (8) yields

$$\begin{aligned} q(x) &= \frac{1}{\tau} \mathbf{w}^T \Phi(x) - (\mathbf{w}^T \Phi_x(x))^T f(x) \\ &\quad + \frac{1}{2} (\mathbf{w}^T \Phi_x(x))^T G(x) R^{-1} G^T(x) (\mathbf{w}^T \Phi_x(x)). \end{aligned} \quad (14)$$

Hence the scalar function $q(x)$ is determined from the solution to (14).

Here we develop a new algorithm for the inverse optimality design which does not require solving the forward problem. Given a control law and the system dynamics, the cost function can be recovered by parameterizing and inferring the optimal value function rather than the cost function, and then computing the cost function with an explicit formula.

3. INVERSE DESIGN FOR STOCHASTIC SYSTEM

3.1 Problem Formulation

Consider the nonlinear dynamical system described by the stochastic differential equations

$$d\mathbf{x} = f(\mathbf{x}) dt + G(\mathbf{x}) (\mathbf{u} dt + \sigma \mathbf{u} d\varepsilon) + \mathcal{F}(\mathbf{x}) d\omega, \quad (15)$$

where $\mathbf{x} \in \mathbb{R}^{n_x}$ is the state variable, $\mathbf{u} \in \mathbb{R}^{n_u}$ is the control input, the random variables $\varepsilon \in \mathbb{R}^{n_\varepsilon}, \omega \in \mathbb{R}^{n_\omega}$ are independent of each other and have multidimensional Gaussian distributions with zero means and identity covariances respectively. Here the control input \mathbf{u} is disturbed by the multiplicative noise ε , where σ is a unitless quantity that defines the noise magnitude relative to the magnitude of the control signal. Thus, the control-dependent and additive noise terms in (15) have covariances $\sigma^2 G(\mathbf{x}) \mathbf{u} \mathbf{u}^T G^T(\mathbf{x})$ and $\mathcal{F}(\mathbf{x}) \mathcal{F}^T(\mathbf{x})$ respectively. And $f(\mathbf{x})$, $G(\mathbf{x})$ and $\mathcal{F}(\mathbf{x})$ are known nonlinearities.

Define the value function (cost-to-go function) $v(t, \mathbf{x})$ as the total cost expected to accumulate if the system is initialized in state \mathbf{x} at time t , and controlled according to the control law

$$v(t, \mathbf{x}) \triangleq E \int_t^\infty e^{-t/\tau} \left[q(\mathbf{x}) + \frac{1}{2} \mathbf{u}^T R \mathbf{u} \right] dt, \quad (16)$$

where $q(\mathbf{x})$ is a smooth function of its arguments, $R \in \mathbb{R}^{n_u \times n_u}$ is symmetric and positive definite, and $\tau > 0$ represents the constant discount factor. The expectation is taken over the instantiations of the stochastic process.

The inverse optimal control problem can now be formulated as follows. Given the dynamic system (15) and a feedback control law \mathbf{u} as

$$\mathbf{u} = \Psi(\mathbf{x}) \quad (17)$$

which can result in an asymptotically stabilized closed-loop system

$$d\mathbf{x} = f(\mathbf{x}) dt + G(\mathbf{x}) \left(\Psi(\mathbf{x}) dt + \sigma \Psi(\mathbf{x}) d\varepsilon \right) + \mathcal{F}(\mathbf{x}) d\omega, \quad (18)$$

the objective here is to estimate the cost function (including $q(\mathbf{x})$ and R) for which the control law (17) is optimal. Here we assume that the optimal control law exists and the necessary and sufficient condition for \mathbf{u} to be optimal can be obtained by solving the HJB equation.

3.2 Cost function Estimation

Let us define the Hamiltonian \mathcal{H} as

$$\begin{aligned} \mathcal{H}(\mathbf{x}, \mathbf{u}, v(t, \mathbf{x})) &= q(\mathbf{x}) + \frac{1}{2} \mathbf{u}^T R \mathbf{u} \\ &\quad + v_x^T \left(f(\mathbf{x}) + G(\mathbf{x}) \mathbf{u} \right) \\ &\quad + \frac{1}{2} \text{tr} \left(\sigma^2 G(\mathbf{x}) \mathbf{u} \mathbf{u}^T G^T(\mathbf{x}) v_{\mathbf{x}\mathbf{x}} \right) \\ &\quad + \frac{1}{2} \text{tr} \left(\mathcal{F}(\mathbf{x}) \mathcal{F}^T(\mathbf{x}) v_{\mathbf{x}\mathbf{x}} \right), \end{aligned} \quad (19)$$

where v_x and $v_{\mathbf{x}\mathbf{x}}$ represent the first and second order partial derivative of $v(t, \mathbf{x})$ over \mathbf{x} respectively.

A necessary and sufficient condition for a control law

$$\mathbf{u}^* = \Psi(\mathbf{x}) = -\bar{R}^{-1} G(\mathbf{x})^T v_x^*, \quad (20)$$

$$\bar{R} \triangleq R + \sigma^2 G^T(\mathbf{x}) v_{\mathbf{x}\mathbf{x}}^* G(\mathbf{x}) \quad (21)$$

to be optimal is that it satisfies the following HJB equation

$$-v_t^*(t, \mathbf{x}) = \min_{\mathbf{u}(t)} \left\{ \mathcal{H}(\mathbf{x}(t), \mathbf{u}(t), v^*(t, \mathbf{x})) \right\}. \quad (22)$$

Note that the optimal value function (optimal cost-to-go) $v^*(t, \mathbf{x})$ is the accumulated expected cost if the system is initialized in state \mathbf{x} at time t , and the optimal control law \mathbf{u}^* is applied until the end of optimal control history; $v_t^*(t, \mathbf{x})$ is the derivative of $v^*(t, \mathbf{x})$ over t .

The Hamiltonian (19) can be minimized analytically with respect to \mathbf{u} by solving $\frac{\partial \mathcal{H}}{\partial \mathbf{u}} = 0$, the minimum is achieved at (20) and the result is

$$\begin{aligned} \min_{\mathbf{u}(t)} \left\{ \mathcal{H}(\mathbf{x}, \mathbf{u}, v^*(t, \mathbf{x})) \right\} &= q(\mathbf{x}) + v_x^{*T} f(\mathbf{x}) \\ &\quad - \frac{1}{2} v_x^{*T} G(\mathbf{x}) \bar{R}^{-1} G^T(\mathbf{x}) v_x^* \\ &\quad + \frac{1}{2} \text{tr} \left(v_{\mathbf{x}\mathbf{x}} \mathcal{F}(\mathbf{x}) \mathcal{F}^T(\mathbf{x}) \right). \end{aligned} \quad (23)$$

Substituting (23) into (22), we obtain the HJB equation as follows

$$\begin{aligned} \frac{1}{\tau} v^*(t, \mathbf{x}) &= q(\mathbf{x}) + v_{\mathbf{x}}^{*T} f(\mathbf{x}) \\ &\quad - \frac{1}{2} v_{\mathbf{x}}^{*T} G(\mathbf{x}) \bar{R}^{-1} G^T(\mathbf{x}) v_{\mathbf{x}}^* \\ &\quad + \frac{1}{2} \text{tr} \left(v_{\mathbf{xx}} \mathcal{F}(\mathbf{x}) \mathcal{F}^T(\mathbf{x}) \right), \end{aligned} \quad (24)$$

thus the HJB equation becomes the second order quadratic PDE which doesn't involve the minimum operator.

In order to obtain more explicit results on the optimal value function, we use the following function approximator

$$v^*(t, \mathbf{x}) \cong \sum_{i=1}^m w_i \Phi_i(\mathbf{x}) = \mathbf{w}^T \Phi(\mathbf{x}), \quad (25)$$

where

$$\begin{aligned} \Phi_i(\mathbf{x}) &= \exp \left(-\frac{\|\mathbf{x} - c_i\|^2}{r_i^2} \right), \\ \Phi(\mathbf{x}) &= \left(\Phi_1(\mathbf{x}) \cdots \Phi_m(\mathbf{x}) \right)^T, \\ \mathbf{w} &= (w_1 \ w_2 \ \cdots \ w_m)^T, \end{aligned}$$

$\Phi_i(\mathbf{x})$ is the i^{th} Gaussian radial basis function (GRBF), c_i defines its center and r_i effectively determines the width of GRBF, and $\Phi(\mathbf{x}) \in \mathbf{R}^m$ is defined as a vector of GRBFs; w_i is the associated weight of $\Phi_i(\mathbf{x})$, $\mathbf{w} \in \mathbf{R}^m$ is defined as a vector of linear weights.

Taking the derivative over \mathbf{x} on both sides of (25) yields

$$v_{\mathbf{x}}^* = \mathbf{w}^T \Phi_{\mathbf{x}}, \quad (26)$$

$$v_{\mathbf{xx}}^* = \mathbf{w}^T \Phi_{\mathbf{xx}}, \quad (27)$$

where $\Phi_{\mathbf{x}}, \Phi_{\mathbf{xx}}$ defines the first and second derivative of $\Phi(\mathbf{x})$ over \mathbf{x} respectively. Here we suppress the dependence on \mathbf{x} for clarity. Combining equations (20)-(21) and (26)-(27), the linear parameters \mathbf{w} can then be solved by the linear regression method.

By choosing R as a positive definite matrix, the cost function can be recovered by computing

$$\begin{aligned} q(\mathbf{x}) &= \frac{1}{\tau} \mathbf{w}^T \Phi - (\mathbf{w}^T \Phi_{\mathbf{x}})^T f(\mathbf{x}) \\ &\quad + \frac{1}{2} (\mathbf{w}^T \Phi_{\mathbf{xx}})^T G(\mathbf{x}) \bar{R}^{-1} G^T(\mathbf{x}) (\mathbf{w}^T \Phi_{\mathbf{x}}) \\ &\quad - \frac{1}{2} \text{tr} \left(\mathbf{w}^T \Phi_{\mathbf{xx}} \mathcal{F}(\mathbf{x}) \mathcal{F}^T(\mathbf{x}) \right). \end{aligned} \quad (28)$$

Here we extend our results to the stochastic system. Given a stochastic nonlinear dynamic system and a feedback control law optimized with respect to a discounted cost function, we develop a new algorithm to construct an estimated cost function which can recover the original performance criterion for which this control law is optimal.

4. NUMERICAL EXAMPLES

In the following we present several examples to show the applicability of the new techniques presented in this paper.

4.1 Scalar System

Deterministic Case First we will give a simple example to demonstrate the feasibility of the idea proposed in section 2. Consider a scalar system

$$\dot{x} = -x + u, \quad (29)$$

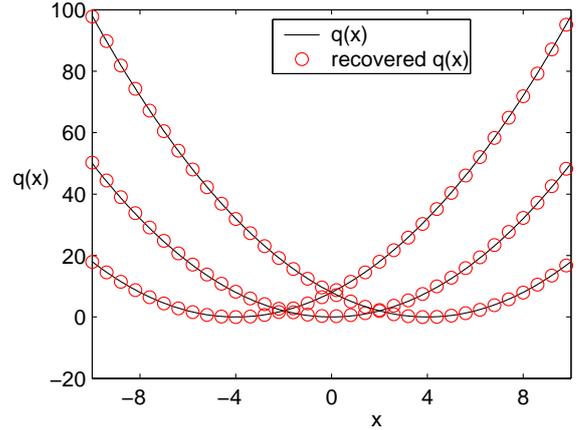


Fig. 1. Cost function $q(x)$ for three different targets: $x_{\text{target}} = -4, 0$ and 4 . (Circle describes the recovered $q(x)$ based on inverse optimality approach, dark line describes the original $q(x)$ function)

and the performance criterion

$$V(x(0)) = \int_0^{\infty} e^{-t/\tau} \left[q(x(t)) + \frac{1}{2} r u(t)^2 \right] dt, \quad r > 0. \quad (30)$$

where the discounted factor $\tau = 5$, and $q(x)$ and r are unknown. Given a control law u under a quadratic cost function where $q(x) = \frac{1}{2}(x - x_{\text{target}})^2$ and $r = 1$, the main objective here is to find $q(x)$ and r for which this control law u is optimal, which means, we would like to know whether the estimated $q(x)$ can recover the original quadratic performance.

Figure 1 illustrates the cost function $q(x)$ for three different targets: $x_{\text{target}} = -4, 0$ and 4 . Black line describes the original function, circle describes the recovered $q(x)$ based on inverse optimal control approach. The simulation demonstrates that, by changing the targets, the cost function can still be recovered very well.

Stochastic Case Consider a stochastic scalar system

$$\dot{x} = -x + u(1 + \sigma \varepsilon), \quad (31)$$

and the performance criterion

$$V(x(0)) = \int_0^{\infty} e^{-t/\tau} \left[q(x(t)) + \frac{1}{2} r u(t)^2 \right] dt, \quad r > 0. \quad (32)$$

where the discounted factor $\tau = 10$, and $q(x)$ and r are unknown. Here the control input u is disturbed by the multiplicative noise, whose standard deviation is 50% of the magnitude of control signal — which means $\sigma = 0.5$ in (31), while ε is a zero-mean Gaussian white noise with unity covariance. Given a control law u under a quadratic cost function where $q(x) = \frac{1}{2}x^2$ and $r = 1$, the main objective here is to find $q(x)$ and r for which this control law u is optimal, which means, we would like to know whether the estimated $q(x)$ can recover the original quadratic performance.

Figure 2 illustrates the performance of the inverse optimality approach we developed in section 3. Black curve corresponds to the original cost function $q(x)$, red circle describes the recovered $q(x)$ based on our approach. The

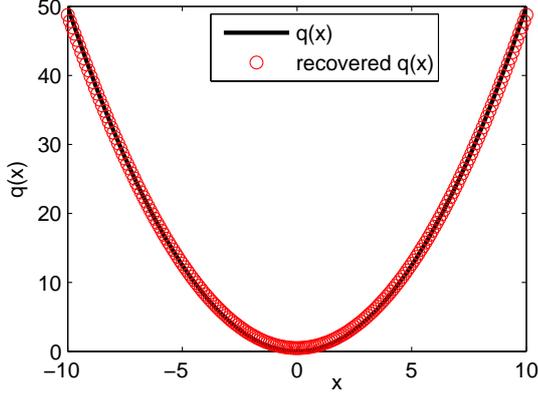


Fig. 2. Cost function $q(x)$. (Circle describes the recovered $q(x)$ based on inverse optimality approach, dark line describes the original $q(x)$ function)

simulation demonstrates that the cost function can be recovered almost exactly.

4.2 Biological Movement System

Consider an arm model with 2 joints (shoulder and elbow), moving in the horizontal plane (Fig. 3). The nonlinear dynamics of this 2-link arm can be represented using the motion equation in the form

$$\mathcal{M}(\theta)\ddot{\theta} + \mathcal{C}(\theta, \dot{\theta}) + \mathcal{B}\dot{\theta} = \tau, \quad (33)$$

where $\theta = [\theta_1 \ \theta_2]^T \in R^2$ is the joint angle vector (shoulder: θ_1 , elbow: θ_2), $\mathcal{M}(\theta) \in R^{2 \times 2}$ is a positive definite symmetric inertia matrix, $\mathcal{C}(\theta, \dot{\theta}) \in R^2$ is a vector centripetal and Coriolis forces, $\mathcal{B} \in R^{2 \times 2}$ is the joint friction matrix, and $\tau \in R^2$ is the joint torque. Here we consider direct torque control where τ is the control signal. In (33), the expressions of the different variables and parameters are given by

$$\mathcal{M} = \begin{pmatrix} a_1 + 2a_2 \cos \theta_2 & a_3 + a_2 \cos \theta_2 \\ a_3 + a_2 \cos \theta_2 & a_3 \end{pmatrix}, \quad (34)$$

$$\mathcal{C} = \begin{pmatrix} -\dot{\theta}_2(2\dot{\theta}_1 + \dot{\theta}_2) \\ \dot{\theta}_1^2 \end{pmatrix} a_2 \sin \theta_2, \quad (35)$$

$$\mathcal{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}, \quad (36)$$

$$a_1 = I_1 + I_2 + m_2 l_1^2, \quad (37)$$

$$a_2 = m_2 l_1 s_2, \quad (38)$$

$$a_3 = I_2, \quad (39)$$

where $b_{11} = b_{22} = 0.05$, $b_{12} = b_{21} = 0.025$, m_i is the mass (1.4kg, 1kg), l_i is the length of link i (30cm, 33cm), s_i is the distance from the joint center to the center of the mass for link i (11cm, 16cm), and I_i is the moment of inertia (0.025kgm^2 , 0.045kgm^2).

Based on equation (33)-(39), we can compute the forward dynamics

$$\ddot{\theta} = \mathcal{M}(\theta)^{-1}(\tau - \mathcal{C}(\theta, \dot{\theta}) - \mathcal{B}\dot{\theta}), \quad (40)$$

and write the system (40) into a state space form

$$\dot{x} = F(x) + G(x)(1 + \sigma_u \varepsilon)u, \quad (41)$$

where the state $x = (\theta_1 \ \theta_2 \ \dot{\theta}_1 \ \dot{\theta}_2)^T$ is four dimensional, the control is given by $u = \tau = (\tau_1 \ \tau_2)^T$. The control input u

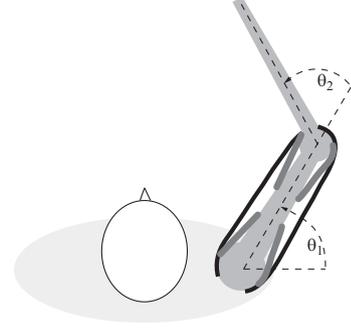


Fig. 3. 2-link arm model (shoulder: θ_1 , elbow: θ_2)

is disturbed by the multiplicative noise, whose standard deviation is 20% of the magnitude of control signal — which means $\sigma_u = 0.2$ in (41), while ε is a zero-mean Gaussian white noise with unity covariance.

The task we are interested in studying is reaching movement, where the arm has to start at an initial position and move to a target in a specified time interval. It also has to stop at the target, and do all that with minimal energy consumption. There are good reasons to believe that such costs are indeed relevant to the neural control of movement Todorov et al. (2002). The cost function for reaching task will depend on the vector of joint angles $[\theta_1 \ \theta_2]^T$ only through the forward kinematic function $e(\theta)$, where e and \dot{e} are the 2D vectors of Cartesian hand coordinates computed as follows

$$e(\theta) = \begin{pmatrix} l_1 \cos \theta_1 + l_2 \cos(\theta_1 + \theta_2) \\ l_1 \sin \theta_1 + l_2 \sin(\theta_1 + \theta_2) \end{pmatrix}, \quad (42)$$

$$\dot{e}(\theta, \dot{\theta}) = \Gamma \begin{pmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \end{pmatrix}, \quad (43)$$

$$\Gamma = \begin{pmatrix} -l_1 \sin \theta_1 - l_2 \sin(\theta_1 + \theta_2) & -l_2 \sin(\theta_1 + \theta_2) \\ l_1 \cos \theta_1 + l_2 \cos(\theta_1 + \theta_2) & l_2 \cos(\theta_1 + \theta_2) \end{pmatrix}, \quad (44)$$

Consider the system (41) and the performance criterion

$$V(x(0)) = \int_0^\infty e^{-t/\tau} \left[q(x(t)) + \frac{1}{2} r u(t)^2 \right] ds, \quad (45)$$

where the discounted factor $\tau = 10$, and $q(x)$ and $r > 0$ are unknown. Given a control law u (Todorov et al. (2005)) under a quadratic cost function

$$q(x(t)) = \|e(\theta(t)) - e^*\|^2 + 0.008 \|\dot{e}(\theta(t), \dot{\theta}(t))\|^2, \quad (46)$$

and $r = 0.00001$, where e^* is the desired target position defined in Cartesian hand coordinates, the second term $\|\dot{e}(\theta(t), \dot{\theta}(t))\|^2$ enforces stopping at the target, the main objective here is to find $q(x)$ and r for which this control law u is optimal.

In order to demonstrate the effectiveness of our design, we now apply the inverse optimality approach developed in section 3 to the reaching movement for the 2-link arm model described above. Given a sequence of control law (joint torques) which can move the arm from the starting position (shown as star in Fig 4A) to the target (shown as circle in Fig 4A), where Fig 4A shows the hand trajectory and Fig 4B shows the velocity profile, the simulation result shown in Fig 5 demonstrates that the

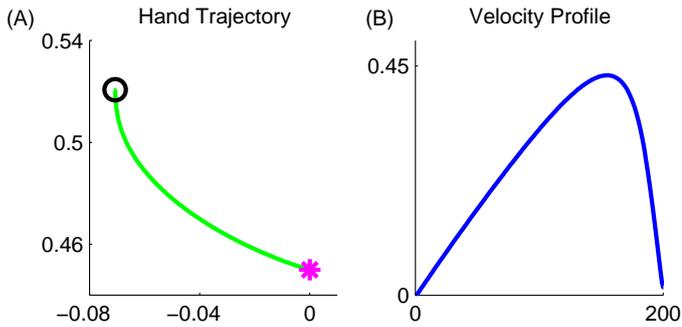


Fig. 4. (A) Hand path for movement in Cartesian hand space; (B) Speed profile

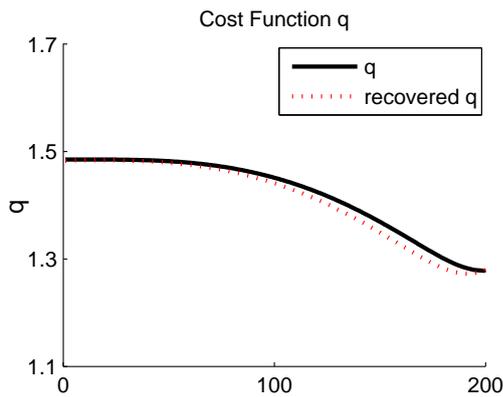


Fig. 5. Cost function $q(x)$. (Red dotted line describes the recovered $q(x)$ based on inverse optimality approach, dark line describes the original $q(x)$ function)

cost function can be recovered very well (relative error $< 2\%$).

5. CONCLUSION AND DISCUSSION

This paper developed a new algorithm for inverse optimality design applicable to deterministic and stochastic nonlinear dynamical systems. The efficiency of the algorithm is the insight of recovering the value function. Assuming a nonlinear stochastic system is given, and a control law under a non-quadratic cost function has already obtained, the new method demonstrates that the estimated cost function can recover the original performance for which this control law is optimal. Although our work is motivated by studying biological movement control, the present results could be of interest to a wider audience. The most important is that our proposed method used the general HJB equation which can deal with nonlinear dynamic cases, and our approximation to cost function uses Gaussian basis functions which do not assume a quadratic form.

When we apply the inverse optimality approach to the biological movement system, there are also a few interesting questions: 1) Considering two different biological models, one is the realistic arm model driven by the muscles, the other is the artificial arm model driven by the electronic motor, we ask them to make the same movement, such as reaching a target, if we obtain the same optimal trajectory, can we recover the same cost function? 2) When we make the reaching movement which is disturbed by a robot, we

know that the brain will learn to adapt this disturbance. How could we deal with this kind of inverse optimality problem? All these open questions are deserved to be solved and will be very beneficial to understand the reality how the brain controls and implements the movement.

REFERENCES

- Anderson, B.D.O and Moore, J.B. (1990). *Optimal Control: Linear Quadratic Methods*. Prentice Hall, Englewood Cliff, NJ.
- Deng, H. and Krstic, M. (1997). Stochastic nonlinear stabilization Part II: Inverse optimality. *Systems and Control Letters*, Vol.32, 151–159.
- Deng, H. and Krstic, M. (1999). Output-feedback stochastic nonlinear stabilization. *IEEE Transactions on Automatic Control*, Vol.44, 328–333.
- Dvijotham, K. and Todorov, E. (2010). Inverse optimal control with linearly-solvable MDPs. *Proceedings of the 27th International Conference on Machine Learning*.
- Freeman, R.A. and Kokotović, P. (1996). Inverse optimality in robust stabilization. *SIAM Journal on Control and Optimization*, Vol.34, Issue 4, 1365–1391.
- Kalman, R.E. (1964). When is a linear control system optimal? *Journal of Basic Engineering*, 51–60.
- Krstic, M. and Tsiotras, P. (1999). Inverse optimal stabilization of a rigid spacecraft. *IEEE Transactions on Automatic Control*, Vol.44, 1042–1049.
- Lewis, F.L. and Syrmos, V.L. (1995). *Optimal Control*. John Wiley and Sons, New York, USA.
- Li, W. and Todorov, E. (2007). Iterative linearization methods for approximately optimal control and estimation of non-linear stochastic system. *International Journal of Control*, Vol.80, Issue 9, 1439–1453.
- Li, Z.H. and Krstic, M. (1997). Optimal design of adaptive tracking controllers for non-linear systems. *Automatica*, Vol.33, No.8, 1459–1473.
- Moylan, P.J. and Anderson, B.D. (1973). Nonlinear regulator theory and an inverse optimal control problem. *IEEE Transactions on Automatic Control*, Vol.AC-18, No.5, 460–465.
- Nedic, A. and Bertsekas, D.P. (2003). Least-squares policy evaluation algorithms with linear function approximation. *Journal of Discrete Event Systems*, Vol.13, 79–110.
- Sepulchre, R., Janković, M., and Kokotović, P. (1997). *Constructive Nonlinear Control*. Springer.
- Tang, C. and Basar, T. (2002). Inverse optimal controller design for strict-feedback stochastic systems with exponential-of-integral cost. *Proceeding of the 15th IFAC World Congress*, Barcelona, Spain.
- Thau, F.E. (1967). On the inverse optimal control problem for a class of nonlinear autonomous systems. *IEEE Transactions on Automatic Control*, Vol.AC-12, No.6, 674–681.
- Todorov, E. and Jordan, M. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, Vol.5, No.11, 1226–1235.
- Todorov, E. and Li, W. (2005). A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems. *Proceedings of 2005 American Control Conference*, 300–306.
- Yokoyama, R. and Kinnen, E. (1972). The inverse problem of the optimal regulator. *IEEE Transactions on Automatic Control*, Vol.AC-17, No.4, 497–504.