In Proceedings of the 45th IEEE Conference on Decision and Control (2006)

# An Iterative Optimal Control and Estimation Design for Nonlinear Stochastic System

Weiwei Li[†] and Emanuel Todorov[‡]

***Abstract*— This paper presents an iterative Linear-Quadratic-Gaussian method for locally-optimal control and estimation of nonlinear stochastic systems. The new method constructs an affine feedback control law obtained by minimizing a novel quadratic approximation to the optimal cost-to-go function. It also constructs a non-adaptive filter optimized with respect to the current control law. The control law and filter are iteratively improved until convergence. The performance of the algorithm is illustrated on a complex biomechanical control problem involving a stochastic model of the human arm.**

## I. Introduction

Optimal control of partially-observable stochastic systems is generally intractable, because the optimal estimator tends to be infinite-dimensional and consequently the optimal controller is also infinite-dimensional. The only notable exception is the Linear-Quadratic-Gaussian (LQG) setting [1], [2], [4], [5], [6], [9], where the posterior probability over the system state is Gaussian and the optimal controller only depends on the mean of that Gaussian. Unfortunately many real-world problems (including the biological control problems we are interested in) are not LQG and are not amenable to global LQG approximations. Local LQG approximations, on the other hand, are often quite reasonable: they rely on local low-order polynomial approximations to the system dynamics, cost function and noise log-probability — all of which tend to be smooth functions. It then makes sense to construct approximately-optimal estimators and controllers by starting with a local LQG approximation, finding the optimal solution, using it to construct a new LQG approximation, and iterating until convergence.

This idea is reminiscent of second-order (Newton) methods for numerical optimization: the function that needs to be optimized is rarely quadratic, but nevertheless one can approximate it locally with a quadratic, find the minimum, construct a new quadratic approximation around that minimum, and iterate. The reason for choosing quadratics is very simple: they represent the most complex numerical optimization problem we know how to solve analytically. Similarly, LQG is the most complex optimal control problem we know how to solve analytically. Thus iterative LQG methods have the potential to play the role that second-order methods have played in numerical optimization (namely, they have become the method of choice). For some reason that potential has not been fully realized. The goal of the present paper is to fill this gap, by iteratively constructing extended Kalman filter and linear-quadratic regulator adapted to the approximation setting.

This work represents the convergence of two lines of research we have previously pursued. In one line of research, we derived an iterative algorithm for optimal estimation and control of partially-observable linear-quadratic systems subject to state-dependent and control-dependent noise [8]. This was possible due to a restriction to non-adaptive filters. In another line of research, we derived an iterative algorithm for optimal control of fully-observable stochastic nonlinear systems with arbitrary costs and control constraints [3], [7]. This was possible due to a novel approximation to the optimal cost-to-go function. Here we combine these ideas, and derive an algorithm which can handle partially-observable nonlinear systems, non-quadratic costs, state-dependent and control-dependent noise, and control constraints. In section II we formalize the problem we want to solve. In sections III and IV we present an LQG approximation to that problem, and compute an approximately-optimal control law under the assumption that state estimates are obtained by an unbiased non-adaptive linear filter. In section V we derive the optimal filter corresponding to the given control law. The two results together provide an iterative coordinate-descent algorithm, which is guaranteed to converge to a filter and a control law optimal with respect to each other. In section VI we illustrate the application of our method and explore its convergence properties, in the context of reaching movements and obstacle avoidance movements using a model of the human arm.

## II. Problem Formulation

Consider the nonlinear dynamical system described by the stochastic differential equations

$$d\mathbf{x}^p = f\big(\mathbf{x}^p, \mathbf{u}^p\big)dt + \mathcal{F}\big(\mathbf{x}^p, \mathbf{u}^p\big)\, d\omega(t), \tag{1}$$

along with the output equation

$$d\mathbf{y}^p(t) = g\big(\mathbf{x}^p, \mathbf{u}^p\big)dt + \mathcal{G}\big(\mathbf{x}^p, \mathbf{u}^p\big)\, dv(t), \tag{2}$$

where state variable $\mathbf{x}^p \in \mathbb{R}^{n_x}$, control input $\mathbf{u}^p \in \mathbb{R}^{n_u}$, measurement output $\mathbf{y}^p \in \mathbb{R}^{n_y}$, and standard Brownian motion noise $\omega \in \mathbb{R}^{n_\omega}$, $v \in \mathbb{R}^{n_v}$ are independent of each other. We define the cost-to-go function $v^\pi(t, \mathbf{x}^p)$ as the total cost expected to accumulate if the system is initialized in state $\mathbf{x}^p$ at time $t$, and controlled until time $T$ according to the control law $\mathbf{u}^p$. The admissible control signals may be constrained: $\mathbf{u}^p \in \mathcal{U}$.

This work is supported by NSF Grant ECS-0524761.

[†]Department of Mechanical and Aerospace Engineering, University of California San Diego, La Jolla, CA 92093-0411 wwli@mechanics.ucsd.edu

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

The objective of optimal control is to find the optimal control law $\mathbf{u}^*$ that minimizes $v^\pi(0, \mathbf{x}^p(0))$. Note that the globally-optimal control law does not depend on a specific initial state. However, finding this control law in complex problems is unlikely. Instead, we seek locally-optimal control laws: we will present an LQG approximation to our original optimal control problem and compute an approximately-optimal control law. The present formulation in this paper assumes that the state of system is measurable through delayed and noisy sensors, therefore we will also design an optimal filter in order to extract the accurate state information from noisy measurement data.

## III. Local LQG Approximation

### A. Linearization

In this paper the locally-optimal control law is computed using the method of dynamic programming. Time is discretized as $k = 1, \cdots, N$, with time step $\Delta t = T/(N-1)$. Our derived algorithm is iterative. Each iteration starts with a nominal control sequence $\bar{\mathbf{u}}_k^p$, and a corresponding nominal trajectory $\bar{\mathbf{x}}_k^p$ obtained by applying $\bar{\mathbf{u}}_k^p$ to the deterministic system $\dot{\mathbf{x}}^p = f(\mathbf{x}^p, \mathbf{u}^p)$ with $\bar{\mathbf{x}}^p(0) = \mathbf{x}_0^p$. This can be done by Euler integration $\bar{\mathbf{x}}_{k+1}^p = \bar{\mathbf{x}}_k^p + \Delta t\ f(\bar{\mathbf{x}}_k^p, \bar{\mathbf{u}}_k^p)$.

By linearizing the system dynamics and quadratizing the cost functions around $(\bar{\mathbf{x}}_k^p, \bar{\mathbf{u}}_k^p)$, we obtain a discrete-time linear dynamical system with quadratic cost. Note that the linearized dynamics no longer describe the state and control variables, instead they describe the state and control deviations $\mathbf{x}_k = \mathbf{x}_k^p - \bar{\mathbf{x}}_k^p$, $\mathbf{u}_k = \mathbf{u}_k^p - \bar{\mathbf{u}}_k^p$, and $\mathbf{y}_k = \mathbf{y}_k^p - \bar{\mathbf{y}}_k^p$, where the value of the outputs at the operating point are defined as $\bar{\mathbf{y}}_k^p = g(\bar{\mathbf{x}}_k^p, \bar{\mathbf{u}}_k^p)$. Written in terms of these deviations — state variable $\mathbf{x}_k \in \mathbb{R}^{n_x}$, control input $\mathbf{u}_k \in \mathbb{R}^{n_u}$, measurement output $\mathbf{y}_k \in \mathbb{R}^{n_y}$, the modified LQG approximation to our original optimal control problem becomes

$$\mathbf{x}_{k+1} = A_k \mathbf{x}_k + B_k \mathbf{u}_k + \mathcal{C}_k(\mathbf{x}_k, \mathbf{u}_k)\xi_k, \tag{3}$$

$$\mathbf{y}_k = F_k \mathbf{x}_k + E_k \mathbf{u}_k + \mathcal{D}_k(\mathbf{x}_k, \mathbf{u}_k)\eta_k, \tag{4}$$

$$\begin{aligned}\text{cost}_k = q_k + \mathbf{x}_k^T \mathbf{q}_k + \frac{1}{2}\mathbf{x}_k^T Q_k \mathbf{x}_k \\ + \mathbf{u}_k^T \mathbf{r}_k + \frac{1}{2}\mathbf{u}_k^T R_k \mathbf{u}_k + \mathbf{u}_k^T P_k \mathbf{x}_k,\end{aligned} \tag{5}$$

where

$$A_k = I + \Delta t\ \frac{\partial f}{\partial \mathbf{x}_k^p}, \qquad B_k = \Delta t\ \frac{\partial f}{\partial \mathbf{u}_k^p}, \tag{6}$$

$$F_k = \frac{\partial g}{\partial \mathbf{x}_k^p}, \qquad E_k = \frac{\partial g}{\partial \mathbf{u}_k^p}, \tag{7}$$

$$\begin{aligned}\mathcal{C}_k(\mathbf{x}_k, \mathbf{u}_k) \triangleq \Big[\mathbf{c}_{1,k} + C_{1,k}^{\mathbf{x}} \mathbf{x}_k + C_{1,k}^{\mathbf{u}} \mathbf{u}_k,\ \cdots\ , \\ \mathbf{c}_{n_\omega,k} + C_{n_\omega,k}^{\mathbf{x}} \mathbf{x}_k + C_{n_\omega,k}^{\mathbf{u}} \mathbf{u}_k\Big],\end{aligned} \tag{8}$$

$$\begin{aligned}\mathcal{D}_k(\mathbf{x}_k, \mathbf{u}_k) \triangleq \Big[\mathbf{d}_{1,k} + D_{1,k}^{\mathbf{x}} \mathbf{x}_k + D_{1,k}^{\mathbf{u}} \mathbf{u}_k,\ \cdots\ , \\ \mathbf{d}_{n_v,k} + D_{n_v,k}^{\mathbf{x}} \mathbf{x}_k + D_{n_v,k}^{\mathbf{u}} \mathbf{u}_k\Big],\end{aligned} \tag{9}$$

$$\mathbf{c}_{i,k} = \sqrt{\Delta t}\ \mathcal{F}^{[i]}, \qquad C_{i,k}^{\mathbf{x}} = \sqrt{\Delta t}\ \frac{\partial \mathcal{F}^{[i]}}{\partial \mathbf{x}_k^p},$$

$$C_{i,k}^{\mathbf{u}} = \sqrt{\Delta t}\ \frac{\partial \mathcal{F}^{[i]}}{\partial \mathbf{u}_k^p}, \qquad \mathbf{d}_{i,k} = \frac{1}{\sqrt{\Delta t}}\ \mathcal{G}^{[i]},$$

$$D_{i,k}^{\mathbf{x}} = \frac{1}{\sqrt{\Delta t}}\ \frac{\partial \mathcal{G}^{[i]}}{\partial \mathbf{x}_k^p}, \qquad D_{i,k}^{\mathbf{u}} = \frac{1}{\sqrt{\Delta t}}\ \frac{\partial \mathcal{G}^{[i]}}{\partial \mathbf{u}_k^p},$$

and

$$q_k = \Delta t\ \ell, \quad \mathbf{q}_k = \Delta t\ \frac{\partial \ell}{\partial \mathbf{x}_k^p}, \quad Q_k = \Delta t\ \frac{\partial^2 \ell}{\partial (\mathbf{x}_k^p)^2},$$

$$\mathbf{r}_k = \Delta t\ \frac{\partial \ell}{\partial \mathbf{u}_k^p}, \quad R_k = \Delta t\ \frac{\partial^2 \ell}{\partial (\mathbf{u}_k^p)^2}, \quad P_k = \Delta t\ \frac{\partial^2 \ell}{\partial \mathbf{u}_k^p \partial \mathbf{x}_k^p},$$

are computed at each $(\bar{\mathbf{x}}_k^p, \bar{\mathbf{u}}_k^p)$.

The initial state has known mean $\hat{\mathbf{x}}_1$ and covariance $\Sigma_1$. All the matrices $A_k, B_k, F_k, E_k, \mathbf{c}_{i,k}, C_{i,k}^{\mathbf{x}}, C_{i,k}^{\mathbf{u}}, \mathbf{d}_{j,k}, D_{j,k}^{\mathbf{x}}, D_{j,k}^{\mathbf{u}}$ $(i = 1, \cdots, n_\omega, \text{and } j = 1, \cdots, n_v)$ are assumed to be given with the proper dimensions. The independent random variables $\xi_k \in \mathbb{R}^{n_\omega}$ and $\eta_k \in \mathbb{R}^{n_v}$ are zero-mean Gaussian white noises with covariances $\Omega^\xi = I$ and $\Omega^\eta = I$ respectively. Note that $\mathcal{F}^{[i]}$ and $\mathcal{G}^{[i]}$ denote the $i^{th}$ column of matrix $\mathcal{F} \in \mathbb{R}^{n_x \times n_\omega}$ and $\mathcal{G} \in \mathbb{R}^{n_y \times n_v}$ respectively. At the final time step $k = N$, the cost is defined as $q_N + \mathbf{x}_N^T \mathbf{q}_N + \frac{1}{2}\mathbf{x}_N^T Q_N \mathbf{x}_N$, where $q_N = h, \mathbf{q}_N = \frac{\partial h}{\partial \mathbf{x}_N^p}$, and $Q_N = \frac{\partial^2 h}{\partial (\mathbf{x}_N^p)^2}$.

Here we are using a noise model which includes control-dependent, state-dependent and additive noises. This is sufficient to capture noise in the system — which is what we are mainly interested in. Considering the sensorimotor control, noise in the motor output increases with the magnitude of the control signal. Incorporating the state-dependent noise in analysis of sensorimotor control could allow more accurate modelling of feedback form sensory modalities and various experimental perturbations. In the study of estimation and control design for the system with control-dependent and state-dependent noises, the well-known Separation Principle of standard LQG design is violated. This complicates the problem substantially, and forces us to develop a new structure of recursive controller and estimator.

### B. Computing the cost-to-go function

In practical situations, the state of the controlled plant are only available through noisy measurement. While the implementation of the optimal control law depends on the state of the system, we have to design an estimator in order to extract the correct information of the state. In this paper we are assuming that the approximately-optimal control law is allowed to be an affine function of $\hat{\mathbf{x}}_k$ — the unbiased estimate of state $\mathbf{x}_k$, and the estimator has the form

$$\hat{\mathbf{x}}_{k+1} = A_k \hat{\mathbf{x}}_k + B_k \mathbf{u}_k + K_k(\mathbf{y}_k - F_k \hat{\mathbf{x}}_k - E_k \mathbf{u}_k), \tag{10}$$

where the filter gains $K_k$ are non-adaptive, i.e. they are determined in advance and cannot change as a function of the specific controls and observations within a simulation run. The detailed derivation for computing the filter gain $K_k$ will be presented in section V.

The approximately-optimal control law for the LQG approximation will be shown to be affine, in the form

$$\mathbf{u}_k = \boldsymbol{\pi}_k(\hat{\mathbf{x}}_k) = l_k + L_k\hat{\mathbf{x}}_k, \qquad k = 1, \cdots, N, \tag{11}$$

where $l_k$ describes the open-loop control component (it arises because we are dealing with state and control deviations, and is needed to make the algorithm iterative), $L_k$ is the feedback control gain. The control law we design is approximately-optimal because we may have control constraints and non-convex costs, and also because we use linear Gaussian approximations. Let the cost-to-go function $v_k(\mathbf{x}_k, \hat{\mathbf{x}}_k)$ be the total cost expected to accumulate if the system (3) is initialized in state $\mathbf{x}_k$ at time step $k$, and controlled according to $\boldsymbol{\pi}_k$ for the remaining time steps.

*Lemma 1:* Suppose the control law $\boldsymbol{\pi}_k$ for system (3)-(5) has already been designed for time steps $k+1, \cdots, N$. If the control law is affine in the form (11), then the cost-to-go function $v_k(\mathbf{x}_k, \hat{\mathbf{x}}_k)$ has the form

$$\begin{aligned} v_k(\mathbf{x}_k, \hat{\mathbf{x}}_k) = \frac{1}{2}\mathbf{x}_k^T S_k^{\mathbf{x}} \mathbf{x}_k + \frac{1}{2}\hat{\mathbf{x}}_k^T S_k^{\hat{\mathbf{x}}} \hat{\mathbf{x}}_k + \mathbf{x}_k^T S_k^{\mathbf{x}\hat{\mathbf{x}}} \hat{\mathbf{x}}_k \\ + \mathbf{x}_k^T \mathbf{s}_k^{\mathbf{x}} + \hat{\mathbf{x}}_k^T \mathbf{s}_k^{\hat{\mathbf{x}}} + \mathsf{s}_k \end{aligned} \tag{12}$$

where the parameters $S_k^{\mathbf{x}}, S_k^{\hat{\mathbf{x}}}, S_k^{\mathbf{x}\hat{\mathbf{x}}}, \mathbf{s}_k^{\mathbf{x}}, \mathbf{s}_k^{\hat{\mathbf{x}}}$, *and* $\mathsf{s}_k$ can be computed recursively backwards in time as

$$\begin{aligned} S_k^{\mathbf{x}} &= Q_k + A_k^T S_{k+1}^{\mathbf{x}} A_k + F_k^T K_k^T S_{k+1}^{\hat{\mathbf{x}}} K_k F_k \\ &\quad + 2A_k^T S_{k+1}^{\mathbf{x}\hat{\mathbf{x}}} K_k F_k + \sum_{i=1}^{n_\omega} (C_{i,k}^{\mathbf{x}})^T S_{k+1}^{\mathbf{x}} C_{i,k}^{\mathbf{x}} \\ &\quad + \sum_{i=1}^{n_v} (D_{i,k}^{\mathbf{x}})^T K_k^T S_{k+1}^{\hat{\mathbf{x}}} K_k D_{i,k}^{\mathbf{x}}, \qquad S_N^{\mathbf{x}} = Q_N, \end{aligned} \tag{13}$$

$$\begin{aligned} S_k^{\hat{\mathbf{x}}} &= (A_k - K_k F_k)^T S_{k+1}^{\hat{\mathbf{x}}} (A_k - K_k F_k) + L_k^T H L_k \\ &\quad + L_k^T G^{\hat{\mathbf{x}}} + (G^{\hat{\mathbf{x}}})^T L_k, \qquad S_N^{\hat{\mathbf{x}}} = 0, \end{aligned} \tag{14}$$

$$\begin{aligned} S_k^{\mathbf{x}\hat{\mathbf{x}}} &= F_k^T K_k^T S_{k+1}^{\hat{\mathbf{x}}} (A_k - K_k F_k) \\ &\quad + A_k^T S_{k+1}^{\mathbf{x}\hat{\mathbf{x}}} (A_k - K_k F_k) + (G^{\mathbf{x}})^T L_k, \; S_N^{\mathbf{x}\hat{\mathbf{x}}} = 0, \end{aligned} \tag{15}$$

$$\begin{aligned} \mathbf{s}_k^{\mathbf{x}} &= \mathbf{q}_k + A_k^T \mathbf{s}_{k+1}^{\mathbf{x}} + F_k^T K_k^T \mathbf{s}_{k+1}^{\hat{\mathbf{x}}} + (G^{\mathbf{x}})^T l_k \\ &\quad + \sum_{i=1}^{n_\omega} (C_{i,k}^{\mathbf{x}})^T S_{k+1}^{\mathbf{x}} \mathbf{c}_{i,k} + \sum_{i=1}^{n_v} (D_{i,k}^{\mathbf{x}})^T K_k^T S_{k+1}^{\hat{\mathbf{x}}} K_k \mathbf{d}_{i,k}, \\ &\qquad\qquad \mathbf{s}_N^{\mathbf{x}} = \mathbf{q}_N, \end{aligned} \tag{16}$$

$$\begin{aligned} \mathbf{s}_k^{\hat{\mathbf{x}}} &= (A_k - K_k F_k)^T \mathbf{s}_{k+1}^{\hat{\mathbf{x}}} + L_k^T H l_k + L_k^T \mathbf{g} + (G^{\hat{\mathbf{x}}})^T l_k, \\ &\qquad\qquad \mathbf{s}_N^{\hat{\mathbf{x}}} = 0, \end{aligned} \tag{17}$$

$$\begin{aligned} \mathsf{s}_k &= q_k + \mathsf{s}_{k+1} + l_k^T \mathbf{g} + \frac{1}{2} l_k^T H l_k \\ &\quad + \frac{1}{2}\left( \sum_{i=1}^{n_\omega} \mathbf{c}_{i,k}^T S_{k+1}^{\mathbf{x}} \mathbf{c}_{i,k} + \sum_{i=1}^{n_v} \mathbf{d}_{i,k}^T K_k^T S_{k+1}^{\hat{\mathbf{x}}} K_k \mathbf{d}_{i,k} \right), \\ &\qquad\qquad \mathsf{s}_N = q_N. \end{aligned} \tag{18}$$

and

$$\begin{aligned} H &\triangleq R_k + B_k^T (S_{k+1}^{\mathbf{x}} + S_{k+1}^{\hat{\mathbf{x}}} + 2S_{k+1}^{\mathbf{x}\hat{\mathbf{x}}}) B_k \\ &\quad + \sum_{i=1}^{n_\omega} (C_{i,k}^{\mathbf{u}})^T S_{k+1}^{\mathbf{x}} C_{i,k}^{\mathbf{u}} + \sum_{i=1}^{n_v} (D_{i,k}^{\mathbf{u}})^T K_k^T S_{k+1}^{\hat{\mathbf{x}}} K_k D_{i,k}^{\mathbf{u}}, \\ \mathbf{g} &\triangleq \mathbf{r}_k + B_k^T (\mathbf{s}_{k+1}^{\mathbf{x}} + \mathbf{s}_{k+1}^{\hat{\mathbf{x}}}) + \sum_{i=1}^{n_\omega} (C_{i,k}^{\mathbf{u}})^T S_{k+1}^{\mathbf{x}} \mathbf{c}_{i,k} \\ &\quad + \sum_{i=1}^{n_v} (D_{i,k}^{\mathbf{u}})^T K_k^T S_{k+1}^{\hat{\mathbf{x}}} K_k \mathbf{d}_{i,k}, \\ G^{\mathbf{x}} &\triangleq P_k + B_k^T (S_{k+1}^{\mathbf{x}} + S_{k+1}^{\mathbf{x}\hat{\mathbf{x}}}) A_k \\ &\quad + B_k^T (S_{k+1}^{\hat{\mathbf{x}}} + S_{k+1}^{\mathbf{x}\hat{\mathbf{x}}}) K_k F_k + \sum_{i=1}^{n_\omega} (C_{i,k}^{\mathbf{u}})^T S_{k+1}^{\mathbf{x}} C_{i,k}^{\mathbf{x}} \\ &\quad + \sum_{i=1}^{n_v} (D_{i,k}^{\mathbf{u}})^T K_k^T S_{k+1}^{\hat{\mathbf{x}}} K_k D_{i,k}^{\mathbf{x}}, \\ G^{\hat{\mathbf{x}}} &\triangleq B_k^T (S_{k+1}^{\hat{\mathbf{x}}} + S_{k+1}^{\mathbf{x}\hat{\mathbf{x}}}) (A_k - K_k F_k). \end{aligned}$$

## IV. Optimal controller design

The cost-to-go function $v_k(\mathbf{x}_k, \hat{\mathbf{x}}_k)$ depends on the control $u_k = \boldsymbol{\pi}_k(\hat{\mathbf{x}}_k)$ through the term

$$a(\mathbf{x}_k, \hat{\mathbf{x}}_k, \boldsymbol{\pi}_k) = \frac{1}{2}\boldsymbol{\pi}_k^T H \boldsymbol{\pi}_k + \boldsymbol{\pi}_k^T (\mathbf{g} + G^{\mathbf{x}}\mathbf{x}_k + G^{\hat{\mathbf{x}}}\hat{\mathbf{x}}_k)$$

This expression is quadratic in $\boldsymbol{\pi}_k$ and can be minimized analytically, but the problem is that the minimum depends on $\mathbf{x}_k$ while $\boldsymbol{\pi}_k$ is only a function of $\hat{\mathbf{x}}_k$. To obtain the optimal control law at time step $k$, we have to take an expectation over $\mathbf{x}_k$ conditional on $\hat{\mathbf{x}}_k$, and find the function $\boldsymbol{\pi}_k$ that minimizes the resulting expression. Since $E[\mathbf{x}_k | \hat{\mathbf{x}}_k] = \hat{\mathbf{x}}_k$, we have

$$\begin{aligned} \alpha(\hat{\mathbf{x}}_k, \boldsymbol{\pi}_k) &\triangleq E[a(\mathbf{x}_k, \hat{\mathbf{x}}_k, \boldsymbol{\pi}_k) \mid \hat{\mathbf{x}}_k] \\ &= \frac{1}{2}\boldsymbol{\pi}_k^T H \boldsymbol{\pi}_k + \boldsymbol{\pi}_k^T (\mathbf{g} + G\hat{\mathbf{x}}_k) \end{aligned} \tag{19}$$

where $G = G^{\mathbf{x}} + G^{\hat{\mathbf{x}}}$. Ideally we would choose $\boldsymbol{\pi}_k$ that minimizes $\alpha(\hat{\mathbf{x}}_k, \boldsymbol{\pi}_k)$ subject to whatever control constraints are present. However, this is not always possible within the family of affine control laws $\boldsymbol{\pi}_k(\hat{\mathbf{x}}_k) = l_k + L_k\hat{\mathbf{x}}_k$ that we are considering. Since the goal of the LQG stage is to approximate the optimal controller for the nonlinear system in the vicinity of $\bar{\mathbf{x}}_k^p$, we will give preference to those control laws that are optimal/feasible for small $\mathbf{x}_k$, even if that (unavoidably) makes them sub-optimal/infeasible for larger $\mathbf{x}_k$.

If the symmetric matrix $H$ in (19) is positive semi-definite, we can compute the unconstrained optimal control law

$$\boldsymbol{\pi}_k = -H^{-1}(\mathbf{g} + G\hat{\mathbf{x}}_k), \tag{20}$$

and deal with the control constraints as described below. But when $H$ has negative eigenvalues, there exist $\boldsymbol{\pi}_k's$ that make $a$ arbitrarily negative. Note that the cost-to-go function for the nonlinear problem is always non-negative, but we are using an approximation to the true cost, we may encounter situations where $a$ does not have a minimum. In that case we use $\mathcal{H}$ to resemble $H$, because $H$ still contains correct second-order information; and so the true cost-to-go decreases in the direction $-\mathcal{H}^{-1}(\mathbf{g} + G\hat{\mathbf{x}}_k)$ for any positive definite matrix $\mathcal{H}$. One possibility is to set $\mathcal{H} = H + (\epsilon - \lambda_{min}(H))I$ where $\lambda_{min}(H)$ is the minimum eigenvalue of $H$ and $\epsilon > 0$. This is related to the Levenberg-Marquardt method.

*Lemma 2:* The optimal control law is computed as

$$\begin{aligned}
&\mathbf{u}_k = l_k + L_k\hat{\mathbf{x}}_k, \qquad (21)\\
&l_k = -\mathcal{H}^{-1}\mathbf{g}, \qquad L_k = -\mathcal{H}^{-1}G,\\
&\mathcal{H} = H + \big(\epsilon - \lambda_{min}(H)\big)I, \ \ \epsilon > 0,\\
&H \triangleq R_k + B_k^T(S^{\mathbf{x}}_{k+1} + S^{\hat{\mathbf{x}}}_{k+1} + 2S^{\mathbf{x}\hat{\mathbf{x}}}_{k+1})B_k\\
&\quad + \sum_{i=1}^{n_\omega}(C^{\mathbf{u}}_{i,k})^T S^{\mathbf{x}}_{k+1}C^{\mathbf{u}}_{i,k} + \sum_{i=1}^{n_v}(D^{\mathbf{u}}_{i,k})^T K_k^T S^{\hat{\mathbf{x}}}_{k+1}K_k D^{\mathbf{u}}_{i,k},\\
&\mathbf{g} \triangleq \mathbf{r}_k + B_k^T(\mathbf{s}^{\mathbf{x}}_{k+1} + \mathbf{s}^{\hat{\mathbf{x}}}_{k+1}) + \sum_{i=1}^{n_\omega}(C^{\mathbf{u}}_{i,k})^T S^{\mathbf{x}}_{k+1}\mathbf{c}_{i,k}\\
&\quad + \sum_{i=1}^{n_v}(D^{\mathbf{u}}_{i,k})^T K_k^T S^{\hat{\mathbf{x}}}_{k+1}K_k\mathbf{d}_{i,k},\\
&G \triangleq P_k + B_k^T(S^{\mathbf{x}}_{k+1} + S^{\hat{\mathbf{x}}}_{k+1} + 2S^{\mathbf{x}\hat{\mathbf{x}}}_{k+1})A_k\\
&\quad + \sum_{i=1}^{n_\omega}(C^{\mathbf{u}}_{i,k})^T S^{\mathbf{x}}_{k+1}C^{\mathbf{x}}_{i,k} + \sum_{i=1}^{n_v}(D^{\mathbf{u}}_{i,k})^T K_k^T S^{\hat{\mathbf{x}}}_{k+1}K_k D^{\mathbf{x}}_{i,k},
\end{aligned}$$

where $S^{\mathbf{x}}_{k+1}, S^{\hat{\mathbf{x}}}_{k+1}, S^{\mathbf{x}\hat{\mathbf{x}}}_{k+1}, \mathbf{s}^{\mathbf{x}}_{k+1}, \mathbf{s}^{\hat{\mathbf{x}}}_{k+1}, \mathbf{s}_{k+1}$ can be obtained through (13)-(18) backwards in time.

## V. Optimal estimator design

It is well known that, for models with control-dependent and state-dependent noises, the optimal filter is very difficult to compute in practice. For this kind of models, the construction of suboptimal filters that approximate the optimal one becomes very important.

So far we computed the optimal control law for any fixed sequence of filter gains $K_k$. In order to preserve the optimality of the control law obtained in the previous section and attain an iterative algorithm with guaranteed convergence, we need to compute a fixed sequence of filter gains that are optimal for a given control law. Thus our objective here is the following: given the control law $\mathbf{u}_1, \cdots, \mathbf{u}_{N-1}$ (which is optimal for the previous filter $K_1, \ldots, K_{N-1}$), compute a new suboptimal filter evaluated by minimizing the magnitude of its estimation errors, in conjunction with the given control law, which results in minimal expected cost. Once the iterative algorithm has converged and the control law has been designed, we could use an adaptive filter in place of the fixed-gain filter in run time.

*Lemma 3:* With the definition of the unconditional means $m_k^e \triangleq E[e_k]$, $m_k^{\hat{\mathbf{x}}} \triangleq E[\hat{\mathbf{x}}_k]$, where $e_k$ is the estimation error and $\hat{\mathbf{x}}_k$ is the estimate of the state, and the unconditional covariances $\Sigma_k^e \triangleq E\left[e_k e_k^T\right]$, $\Sigma_k^{\hat{\mathbf{x}}} \triangleq E\left[\hat{\mathbf{x}}_k\hat{\mathbf{x}}_k^T\right]$, and $\Sigma_k^{\hat{\mathbf{x}}e} \triangleq E\left[\hat{\mathbf{x}}_k e_k^T\right]$, the optimal filter gain for system (3)-(5) is computed as

$$\begin{aligned}
&\hat{\mathbf{x}}_{k+1} = A_k\hat{\mathbf{x}}_k + B_k\boldsymbol{\pi}_k + K_k(\mathbf{y}_k - F_k\hat{\mathbf{x}}_k - E_k\boldsymbol{\pi}_k), \qquad (22)\\
&K_k = A_k\Sigma_k^e F_k^T\left(F_k\Sigma_k^e F_k^T + \mathcal{P}_k\right)^{-1}, \qquad (23)\\
&m^{\hat{\mathbf{x}}}_{k+1} = (A_k + B_kL_k)m_k^{\hat{\mathbf{x}}} + K_kF_km_k^e + B_kl_k, \qquad m_1^{\hat{\mathbf{x}}} = \hat{\mathbf{x}}_1 \qquad (24)\\
&m_{k+1}^e = (A_k - K_kF_k)m_k^e, \qquad m_1^e = 0, \qquad (25)\\
&\Sigma^{\hat{\mathbf{x}}}_{k+1} = (A_k + B_kL_k)\Sigma_k^{\hat{\mathbf{x}}}(A_k + B_kL_k)^T + K_kF_k\Sigma_k^eA_k^T\\
&\quad + (A_k + B_kL_k)\Sigma_k^{\hat{\mathbf{x}}e}F_k^TK_k^T\\
&\quad + K_kF_k\Sigma_k^{e\hat{\mathbf{x}}}(A_k + B_kL_k)^T\\
&\quad + \Big((A_k + B_kL_k)m_k^{\hat{\mathbf{x}}} + K_kF_km_k^e\Big)l_k^TB_k^T\\
&\quad + B_kl_k\Big((A_k + B_kL_k)m_k^{\hat{\mathbf{x}}} + K_kF_km_k^e\Big)^T\\
&\quad + B_kl_kl_k^TB_k^T, \qquad \Sigma_1^{\hat{\mathbf{x}}} = \hat{\mathbf{x}}_1\hat{\mathbf{x}}_1^T, \qquad (26)\\
&\Sigma_{k+1}^e = (A_k - K_kF_k)\Sigma_k^eA_k^T + \mathcal{M}_k, \qquad \Sigma_1^e = \Sigma_1, \qquad (27)\\
&\Sigma^{\hat{\mathbf{x}}e}_{k+1} = (A_k + B_kL_k)\Sigma_k^{\hat{\mathbf{x}}e}(A_k - K_kF_k)^T\\
&\quad + B_kl_k(m_k^e)^T(A_k - K_kF_k)^T, \qquad \Sigma_1^{\hat{\mathbf{x}}e} = 0, \qquad (28)\\
&\mathcal{P}_k = E\left[\mathcal{D}_k(\mathbf{x}_k, \boldsymbol{\pi}_k)\ \mathcal{D}_k^T(\mathbf{x}_k, \boldsymbol{\pi}_k)\right], \qquad (29)\\
&\mathcal{M}_k = E\left[\mathcal{C}_k(\mathbf{x}_k, \boldsymbol{\pi}_k)\ \mathcal{C}_k^T(\mathbf{x}_k, \boldsymbol{\pi}_k)\right]. \qquad (30)
\end{aligned}$$

## VI. Numerical Simulations

### A. Application to arm movements

We have thus far tested the algorithm on the reaching movements for a 2-link arm model, which has nonlinear dynamics, non-quadratic costs and multiplicative noise.

*1) 2-link human arm model:* Consider an arm model with 2 joints (shoulder and elbow), moving in the horizontal plane (Fig 1). The inverse dynamics is

$$\mathcal{M}(\theta)\ddot{\theta} + \mathcal{C}(\theta, \dot{\theta}) + \mathcal{B}\dot{\theta} = \tau, \qquad (31)$$

where $\theta \in R^2$ is the joint angle vector (shoulder: $\theta_1$, elbow: $\theta_2$), $\mathcal{M}(\theta) \in R^{2\times 2}$ 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 $\tau$ is the control signal. In (31), the expressions of the different variables and parameters are given by

$$\begin{aligned}
&\mathcal{M} = \begin{pmatrix} a_1 + 2a_2cos\theta_2 & a_3 + a_2cos\theta_2 \\ a_3 + a_2cos\theta_2 & a_3 \end{pmatrix},\\
&\mathcal{C} = \begin{pmatrix} -\dot{\theta}_2(2\dot{\theta}_1 + \dot{\theta}_2) \\ \dot{\theta_1}^2 \end{pmatrix} a_2sin\theta_2, \ \mathcal{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix},\\
&a_1 = I_1 + I_2 + m_2l_1^2, \qquad a_2 = m_2l_1s_2, \qquad a_3 = I_2,
\end{aligned}$$

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

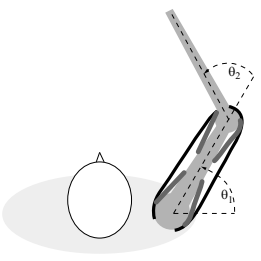



Fig. 1. 2-link arm model.

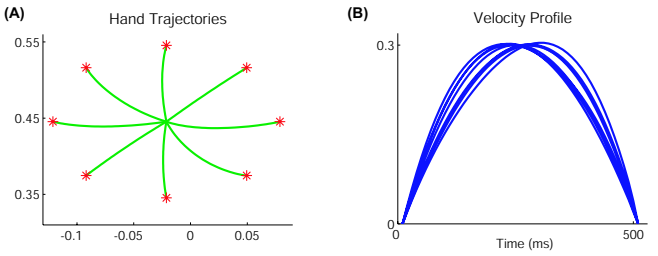



Fig. 2. Fully observable case: average behavior of the ILQG controller for reaching movements, using a 2-link human arm model. (A) Hand paths for movement in 8 directions; (B) Speed profiles.

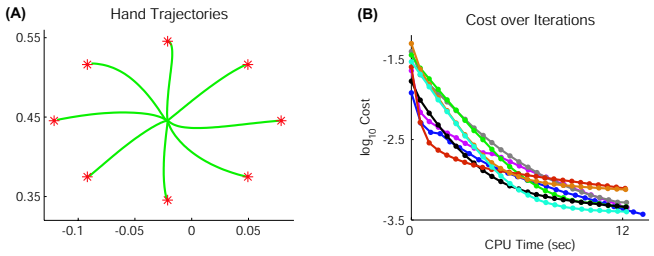



Fig. 3. Partial observable case: average behavior of the ILQG controller and estimator for reaching movements, using a 2-link human arm model. (A) Hand paths for movement in 8 directions; (B) Cost over iterations.

$(0.025kgm^2, 0.045kgm^2)$. Based on equations (31), we can compute the forward dynamics

$$\ddot{\theta} = \mathcal{M}(\theta)^{-1}\big(\tau - \mathcal{C}(\theta,\dot{\theta}) - \mathcal{B}\dot{\theta}\big), \tag{32}$$

and write the system into state space form with $x = (\theta_1\ \theta_2\ \dot{\theta}_1\ \dot{\theta}_2)^T$, $u = \tau = (\tau_1\ \tau_2)^T$. The control input $u$ is disturbed by the multiplicative noise, whose standard deviation is $20\%$ of the magnitude of control signal.

The sensory feedback carries the information about position and velocity

$$y = (\theta_1\ \theta_2\ \dot{\theta}_1\ \dot{\theta}_2)^T + v, \tag{33}$$

where the sensory noise $v$ has zero-mean Gaussian distribution with unity covariance.

*2) Center-out reaching task:* In order to demonstrate the effectiveness of our design, we applied ILQG method to the human arm model described above. The first task we study is reaching movement, where the arm has to start at some 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 [7]. The cost function is defined as

$$J_1 = \Big\|e(\theta(T)) - e^*\Big\|^2 + 0.001\Big\|\dot{e}\Big(\theta(T),\dot{\theta(T)}\Big)\Big\|^2 + \frac{1}{2}\int_0^T 0.0001\|u\|^2\,dt, \tag{34}$$

where $e(\theta)$ and $\dot{e}(\theta,\dot{\theta})$ is the forward kinematics transformation from joint coordinates to end-point coordinates, and the desired target $e^*$ is defined in end-point coordinates. Here we use the center-out reaching task which is commonly studied in the Motor Control — the targets $e^*$ (shown as stars in Fig 2A) are arranged in a circle with 0.1m radius around the starting position. Fig 2 shows average behavior for the fully observable case: hand paths in (A), tangential speed profiles in (B). We found out that the movement kinematics share many features with experimental data on human arm movements.

Now we look at the partial observable case where the states of system are obtained by the estimator. Although the state of the controlled plant are only available through noisy measurement, the movement trajectories shown in Fig 3A illustrates that the hand could still arrive to the desired target position. Another encouraging result is that in terms of CPU time. Fig 3B shows how the total cost decreased over iterations of the algorithm, for reaching in 8 different directions. On average, ILQG found a locally-optimal time-varying feedback control law in about 10 seconds (on a 2.8GHz Pentium 4 machine, in Matlab).

Trajectory-based algorithms related to Pontryagin's Maximum Principle in general find locally-optimal solutions, and complex control problems may exhibit many local minima. Finally, we explored the issue of local minima for the arm control problem. We used 50 different initializations, for each of 8 movement directions. The final trajectories are given in Fig 4, where Fig 4A shows that, for the fully observable case, all the optimization runs converge to a solution very similar to the best solution we find for the corresponding target direction. Fig 4B shows how the cloud of 50 randomly initialized trajectories gradually converge for the partial observable case by using ILQG method. There are local minima, but half the time the algorithm converges to the same result. Therefore, the derived algorithm is relatively very robust, and a small number of restarts of ILQG are sufficient to discover what appears to be the global minimum in a relatively complex control problem.

*3) Reaching task with the obstacle avoidance:* The second task is to implement the reaching task and to avoid the obstacle during the movement, while the obstacle is defined as a circle with a certain radius $r_{obstacle} = 0.02m$, and is arranged in the fixed position. The distance between the starting position and the target is about 0.15m. The arm starts

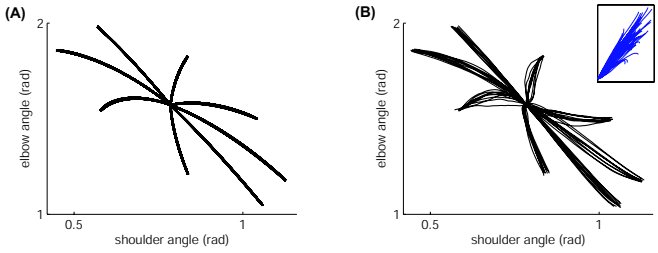



Fig. 4. Hand paths for random 50 initial control laws (blue, inset) and optimized paths (black) to 8 targets obtained by using those initial conditions. (A) fully observable case; (B) partial observable case.

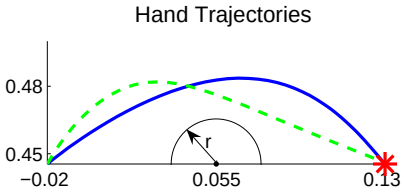
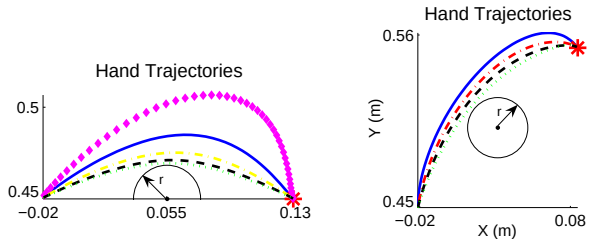



Fig. 5. Average behavior of the ILQG controller and estimator for reaching movement with obstacle avoidance, using a 2-link human arm model. Blue curve: fully observable case; green dashed curve: partial observable case. Note that obstacle circle radius $r = 0.02m$.




(a) contribution of obstacle cost rate (b) contribution of movement duration

Fig. 6. (a) Comparison of movement behavior by choosing different weighting coefficients $k_1$ on the obstacle cost rate (fully observable case). Magenta diamond: $k_1 = 1e-7$; Blue solid: $k_1 = 1e-8$; Yellow dashdot: $k_1 = 1e-9$; Black dashed: $k_1 = 1e-10$; Green dotted: $k_1 = 1e-11$. (b) Comparison of movement behavior by choosing different movement duration (fully observable case). Blue solid: $700msec$; Red dashdot: $500msec$; Black dashed: $350msec$; Green dotted: $200msec$. The obstacle circle radius $r = 0.02m$.

from rest at $\theta_1 = \pi/4, \theta_2 = \pi/2$, and has to reach the specified target and avoid the obstacle during the reaching, with minimal control energy. The cost function is

$$J_2 = \left\| e(\theta(T)) - e^* \right\|^2 + w \left\| \dot{e}\Big(\theta(T), \dot{\theta(T)}\Big) \right\|^2 + \frac{1}{2}\int_0^T r\|u\|^2\, dt + k1 \int_0^T \Big(l(\theta(t)) - r_{obstacle}\Big)^{-2} dt, \tag{35}$$

where the target $e^*$ is defined in end-point coordinates; $l$ is the distance between the hand position and the center of the obstacle center, and $l = \sqrt{\|e(\theta(t)) - e_{ocenter}\|^2}$; the weighting coefficient $w = 0.001, r = 0.0001, k_1 = 1e-8$; and the obstacle radius $r_{obstacle}$ is equal to $0.02m$.

Fig 5 shows the movement trajectories and illustrates how the hand avoid the obstacle circle (marked as dark) and arrive to the desired target position (marked as red star) as close as possible. The blue curve is obtained under the condition that the state variable of the controlled plant is available; while the green dashed curve illustrates the hand movement trajectory tying together the optimal feedback control and estimation design. In the later case, the state of the controlled plant are approximately computed, because of the existence of estimation error, we found out that the green curve becomes more curved at the beginning of the movement compared with the result on fully observable case.

Here the optimal control problem is solved for minimizing a performance criterion. Fig 6(a) shows the behavior of movement trajectories by changing the penalty weighting $k_1$ on the obstacle avoidance in the objective function (35). The bigger the weight penalty $k_1$ is, the hand movement trajectory is further away from the obstacle, and Fig 6(a) exactly explains this phenomena. Another interesting result in behavioral experiments shows that a longer movement duration can be predicted to cause a more curved trajectory. Our simulation result Fig 6(b) quantitatively supported those previous studies with an enormous amount of data.

## VII. Conclusion

This paper developed a new local method for optimal feedback control and estimation design of stochastic nonlinear dynamical systems subject to control constraints. It provided an iterative coordinate-descent algorithm, which is guaranteed to converge to a filter and a control law optimal with respect to each other. 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 approach yields a numerical algorithm with stable convergence achieved through backtracking line search; and convergence in the vicinity of a local minimum is quadratic. Finally, we illustrate its application to reaching movements on a biomechanical model of the human arm.

There are several extensions to the work we presented here. While we assumed a specified final time $T$, the algorithm can be applied in model-predictive mode, using a fixed time horizon rather than a fixed final time. The final cost $h(x)$ will have to be replaced with some approximation to the optimal cost-to-go, but that has to be done whenever fixed-horizon model-predictive control is used.

## References


[1] A. Beghi and D. D'Alessandro, Discrete-time optimal control with control-dependent noise and generalized Riccati difference equations, *Automatica*, 34(8), 1998, pp 1031-1034.

[2] B. Hoff, A computational description of the organization of human reaching and prehension, Ph.D. Thesis, University of Southern California, 1992.

[3] W. Li and E. Todorov, "Iterative linear quadratic regulator design for nonlinear biological movement systems," *In Proceedings of the 1st International Conference on Informatics in Control, Automation and Robotics*, vol. 1, 2004, pp. 222-229.

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

[5] J. B. Moore, X. Y. Zhou and E. B. Lim, Discrete time LQG controls with control dependent noise, *Systems & Control Letters*, 36, 1999, pp 199-206.

[6] M. A. Rami, X. Chen, J. B. Moore and X. Y. Zhou, Solvability and asymptotic behavior of generalized Riccati equations arising in indefinite stochastic LQ controls, *IEEE Transactions on Automatic Control*, 46(3), 2001, pp 428-440.

[7] E. Todorov and W. Li, "Optimal control methods suitable for biomechanical systems," *In proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2003, pp 1758-1761.

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

[9] J. L. Willems and J. C. Willems, Feedback stabilizability for stochastic systems with state and control dependent noise, *Automatica*, 12, 1976, pp 277-283.