Minimum Angular Acceleration Control of Articulated Body Dynamics

Javier R. Movellan
UCSD

***Abstract*— As robots find applications in daily life conditions it becomes important to develop controllers that generate energy efficient movements by restricting variability and utilizing high gains only when necessary. Here we present a computationally light and energy efficient approach (AAC) that combines an anticipatory open-loop controller and a variable gain closed loop controller. The approach is grounded in the theory of stochastic optimal control and feedback linearization. As such it links two important approaches to robot control: (1) the family of Computed Torque Controllers (CTC) that are grounded on feedback linearization and classic feedback control, and (2) a more recent family of controllers that aim at finding approximately optimal trade-offs between task performance and energy consumption. Here we show that AAC controllers are highly energy efficient, when compared to CTC, and exhibit some key properties of human motion.**

I. INTRODUCTION

Computed torque control (CTC) in its various forms [4] is currently the most popular approach to articulated robot control and is arguably responsible for the success of robotics in the assembly line. CTC is light weight, computationally speaking, and can generate precise and repeatable movements. However this precision is achieved at the expense of high torque and energy consumption. This is appropriate for conditions in which variability and uncertainty can be suppressed by controlling the environment and using high energy actuators, like in a high-volume assembly line of identical products. However as robots move to the unconstrained conditions of daily life more sophisticated control approaches become necessary. These new approaches need to generate motion that, while accomplishing the task at hand, consume as least energy as possible. Low energy consumption is important for multiple reasons. First in mobile robots, producing low energy movement is critical for saving battery life. Second, in unconstrained environments high torque movements can be very dangerous to human beings and thus they need to be avoided when they are not required. Third, empirical data suggest that skilled biological motion is fluid and energy efficient [2]. Thus low energy controllers may result on robots with more human-like motion.

Stochastic optimal control provides a powerful formalism to develop low energy robot motion [3], [19], [18]. To this end one simply formulates a criterion function that rewards accomplishment of task goals while it penalizes the use of high torques. Unfortunately robot dynamics are non-linear and exact solutions to the aforementioned optimal control problem are not available. Nevertheless recent years have seen good progress towards the development of numerical methods to find approximately optimal solutions. These include: (1) local iterative approximations like stochastic differential dynamic programming [17] and iLQG [20]; (2) global solutions to the HJB control equation using function approximation methods [15], and (3) approximations based on reinforcement learning methods [1], [14], [16], [8], [10]. While we have seen dramatic progress in this critical area of research, current algorithms are still computationally expensive, and difficult to scale up to robots with a large number of degrees of freedom.

Here we present an approach (AAC) that generates efficient robot controllers, in terms of energy consumption, while being globally optimal, scalable, and computable in real time for large scale robots. The approach is framed on the theory of stochastic optimal control. However rather than using a criterion function that penalizes large torques, it uses a criterion that penalizes large angular accelerations. This results on a remarkable simplification of the control problem so that globally optimal control solutions are possible. While minimizing angular accelerations does not explicitly minimize energy consumption, in practice it results on movements that are energy efficient and that resemble well known properties of biological motion, including the principle of minimum intervention, anticipation, and bell-shaped velocity curves [6], [21], [19].

II. PROBLEM STATEMENT

Consider an articulated body governed by the standard joint angle dynamics

$$M(\theta_t)\ddot{\theta}_t = \tau_t + N(\theta_t, \dot{\theta}_t) + W_t(\theta_t, \dot{\theta}_t, \bar{\theta}_t) \tag{1}$$

where θ_t is the vector of joint angles, $\bar{\theta}_t$ the temporally integrated joint angles, $\dot{\theta}_t$ the angular velocities, and $\ddot{\theta}_t$ the angular accelerations. $M(\theta_t)$ is the moment of inertia matrix, τ_t the vector torques applied by the rotational joint actuators, $N(\theta_t, \dot{\theta}_t)$ is the vector of gravitational, friction and Coriolis/Centripetal forces, and W_t is a zero mean vector of random torques that simulates the effects of uncertainty in the body and environment dynamics.

We are given a set of target pairs of the form $\{(\xi_t, t) : t \in \mathcal{I})$ where ξ_t is a vector of desired joint angles, t the time at which those angles are to be accomplished, and $\mathcal{I}$ a dense index set, e.g., a collection of intervals in the real line. Our goal is to find a closed-loop control law that maximizes the fit between the target joint angles (ξ_t, t) and the obtained joint angles (θ_t, t) while minimizing the angular accelerations needed to achieve these targets.

III. Problem Formalization

We define the performance ρ of a a control policy π as the expected value of the integral, over a fixed time horizon $[t, T]$, of a reward rate R

$$\rho_t(\pi) = E[\int_t^T e^{-\frac{s}{\lambda}} R_s d_s \mid \pi] \tag{2}$$

where $\lambda > 0$ is a scalar controlling the temporal discount for the reward rate. We let the reward rate be the sum of two quadratic criteria: one that penalizes for deviations between the desired and obtained joint angles, and another one that penalizes for large angular accelerations

$$R_t = -(y_t - X_t)' p_t (y_t - X_t) - \ddot{\theta}_t' q_t \ddot{\theta}_t \tag{3}$$

where

$$X_t = \begin{bmatrix} \theta_t \\ \bar{\theta}_t \\ \dot{\theta}_t \end{bmatrix}, \quad y_t = \begin{bmatrix} \xi_t \\ \bar{\xi}_t \\ \dot{\xi}_t \end{bmatrix} \tag{4}$$

$$\tag{5}$$

and p_t, q_t are positive definite matrices. Note under this formulation controllers achieve large reward rates, by having X_t as close as possible to y_t while minimizing the angular acceleration $\ddot{\theta}_t$. For time steps with no task constraints, i.e., for $t \notin \mathcal{I}$ we simply set $p_t = 0$. We model the uncertainty in the environment and body dynamics as Brownian motion using the following Ito-style stochastic differential equation [13]

$$\begin{aligned} dX_t =& aX_t dt + bU_t dt \\ &+ \Big(c_t + \sum_i g_{i,t} X_{i,t} + \sum_j U_{j,t} h_{j,t}\Big) dB_t \end{aligned} \tag{6}$$

where

$$a = \begin{bmatrix} 0 & 0 & I \\ I & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 0 \\ I \end{bmatrix} \tag{7}$$

$$U_t = M^{-1}(\theta_t)\Big(\tau_t + N(\theta_t, \dot{\theta}_t)\Big) = E[\ddot{\theta}_t \mid \tau_t, \theta_t, \dot{\theta}_t] \tag{8}$$

Here c_t, g_t, h_t are dispersion matrices that control the uncertainty in the system dynamics and dB_t is a Brownian motion differential. c_t modulates the effects of noise in a manner that may depend on time. g_t modulates noise in a manner that may depend on joint angles, angular velocities, and integrals of joint angles over time. h_t modulates noise in a manner that may depend on angular accelerations.

IV. Problem Solution

We can use the tools of stochastic optimal control to find the controller π that maximizes the criterion function ρ_t. Note under this formulation the non-linear part of the articulated body dynamics has been moved into U_t, which is the expected angular acceleration. Thus if we let U_t represent the actions of a controller, we find ourselves with a time-dependent linear quadratic tracking problem that can be solved using the standard tools from the theory of stochastic optimal control. The optimal controller found with these tools maps states X_t into desired angular accelerations, i.e

$$U_t = \pi_t(X_t) = E[\ddot{\theta}_t] \tag{9}$$

To map states into torques we simply takes expected values on (1) and solve for the torque τ_t

$$\tau_t(X_t) = M(\theta_t) U_t(X_t) - N(\theta_t, \dot{\theta}_t) \tag{10}$$

While the solution to optimal quadratic tracking is known, for completeness here we sketch its derivation and examine the nature of the resulting controller.

Solutions to stochastic linear quadratic problems of this type are typically found by solving the HJB equation for the value function V of the optimal control policy π

$$V_t(x) = \int_t^T e^{-\frac{s}{\lambda}} R_s ds \mid X_t = x, \pi] \tag{11}$$

We first assume that the value function is quadratic (later we see by induction that the assumption is correct)

$$V_t(x) = -\Big(x' \alpha_t x - 2\beta_t' x + \gamma_t\Big) \tag{12}$$

Under this assumption the HJB equation takes the following form

$$\begin{aligned} x'\dot{\alpha}_t x - 2\dot{\beta}_t' x + \dot{\gamma}_t = \max_u \Big\{ & \frac{1}{\lambda} x' \alpha_t x - \frac{2}{\lambda}\beta_t' x + \frac{1}{\lambda}\gamma_t \\ & - (x - \xi_t)' p_{1,t} (x - \xi_t) + 2p_{2,t}' x \\ & - u'(q_t + \hat{h}_t) u + 2u'(q_t \omega_t + b'(\beta_t - \alpha_t x) - \bar{h}_t - \hat{f}_t x) \\ & - \omega_t' q_t \omega_t + 2(\beta_t - \alpha_t x)'(k_t + ax) \\ & - \mathrm{Tr}[c_t c_t' \alpha_t] - x' \hat{g}_t x - 2\bar{g}_t' x \Big\} \end{aligned} \tag{13}$$

where

$$(\bar{g}_t)_i = \mathrm{Tr}[g_{i,t} c_t' \alpha_t] \tag{14}$$

$$(\hat{g}_t)_{i,j} = \mathrm{Tr}\Big[g_i g_j' \alpha_t\Big] \tag{15}$$

$$(\bar{h}_t)_i = \mathrm{Tr}[h_{i,t} c_t' \alpha_t] \tag{16}$$

$$(\hat{h}_t)_{i,j} = \mathrm{Tr}\Big[h_{i,t} h_{j,t}' \alpha_t\Big] \tag{17}$$

$$(\hat{f}_t)_{i,j} = \mathrm{Tr}\Big[h_{i,t} g_{j,t}' \alpha_t\Big] \tag{18}$$

To find the optimal solution we take the gradient of the RHS of (13) with respect to u and set it to zero

$$-2(q_t + \hat{h}_t)u + 2(q_t \omega_t + b'\beta_t - b'\alpha_t x - \bar{h}_t - \hat{f}_t x) = 0 \tag{19}$$

Thus the optimal policy is a non-linear function of time and an affine function of the extended state x

$$\boxed{\begin{aligned} & u_t = w_{1,t} + w_{2,t} x \\ & w_{1,t} = \bar{q}_t^{-1}(q_t \omega_t + b'\beta_t - \bar{h}_t) \\ & w_{2,t} = -\bar{q}_t (b'\alpha_t + \hat{f}_t) \\ & \bar{q}_t = q_t + \hat{h}_t \end{aligned}} \tag{20}$$

All is left is to find the values for $w_{1,t}$, the open-loop part of the control policy and $w_{2,t}$, the closed loop part. To do so we bring the optimal action back into the HJB equation

$$\begin{aligned} x'\dot{\alpha}_t x - 2\dot{\beta}_t' x + \dot{\gamma}_t &= \frac{1}{\lambda} x'\alpha_t x - \frac{2}{\lambda}\beta_t' x + \frac{1}{\lambda}\gamma_t \\ &- (x-\xi_t)' p_{1,t}(x-\xi_t) + 2p_{2,t}' x \\ &- (w_{1,t} + w_{2,t}x)'\bar{q}(w_{1,t} + w_{2,t}x) \\ &+ 2(w_{1,t} + w_{2,t}x)'\bar{q}(w_{1,t} + w_{2,t}x) \\ &- \omega_t' q_t \omega_t + 2(\beta_t - \alpha_t x)'(k_t + ax) \\ &- \mathrm{Tr}[c_t c_t' \alpha_t] - x'\hat{g}_t x - 2\bar{g}_t' x \end{aligned} \tag{21}$$

Gathering quadratic terms on x we get

$$\dot{\alpha}_t = \frac{1}{\lambda}\alpha_t - p_{1,t} + w_{2,t}'\bar{q}_t w_{2,t} - 2\alpha_t a - \hat{g}_t \tag{22}$$

Gathering linear terms on x we get

$$\begin{aligned} \dot{\beta}_t &= \frac{1}{\lambda}\beta_t - p_{1,t}\xi_t - p_{2,t} - w_{2,t}'\bar{q}_t' w_{1,t} \\ &- a'\beta_t + \alpha_t k_t + \bar{g}_t \end{aligned} \tag{23}$$

Gathering constant terms with respect to x we get

$$\begin{aligned} \dot{\gamma}_t &= \frac{1}{\lambda}\gamma - \xi_t' p_{1,t}\xi_t + w_{1,t}'\bar{q}_t w_{1,t} - \omega_t' q_t \omega_t \\ &+ 2\beta_t' k_t - \mathrm{Tr}[c_t c_t' \alpha_t] \end{aligned} \tag{24}$$

This determines the Ricati equations for the parameters of the optimal value function

$$\boxed{\begin{aligned} &\dot{\alpha}_t = \tfrac{1}{\lambda}\alpha_t - p_{1,t} + w_{2,t}'\bar{q}_t w_{2,t} \\ &\qquad -2\alpha_t a - \hat{g}_t \\[1ex] &\dot{\beta}_t = \tfrac{1}{\lambda}\beta_t - p_{1,t}\xi_t - p_{2,t} - w_{2,t}'\bar{q}_t' w_{1,t} \\ &\qquad -a'\beta_t + \alpha_t k_t + \bar{g}_t \\[1ex] &\dot{\gamma}_t = \tfrac{1}{\lambda}\gamma_t - \xi_t' p_1 \xi_t + w_{1,t}'\bar{q}_t w_{1,t} - \omega_t' q_t \omega_t \\ &\qquad +2\beta_t' k - \mathrm{Tr}[c_t c_t' \alpha_t] \\[1ex] &w_{1,t} = \bar{q}_t^{-1}(q\omega + b'\beta_t - \bar{h}) \\[1ex] &w_{2,t} = -\bar{q}_t^{-1}(b'\alpha_t + \hat{f}) \\[1ex] &\bar{q}_t = q_t + \hat{h}_t \\[1ex] &(\bar{g}_t)_i = \mathrm{Tr}[g_{i,t} c_t' \alpha_t] \\[1ex] &(\hat{g}_t)_{i,j} = \mathrm{Tr}[g_{i,t} g_{j,t}' \alpha_t] \\[1ex] &(\bar{h}_t)_i = \mathrm{Tr}[h_{i,t} c_t' \alpha_t] \\[1ex] &(\hat{h}_t)_{i,j} = \mathrm{Tr}[h_{i,t} h_{j,t}' \alpha_t] \\[1ex] &(\hat{f}_t)_{i,j} = \mathrm{Tr}[h_{i,t} g_{j,t}' \alpha_t] \end{aligned}}$$

with the following terminal conditions

$$\alpha_T = p_{1,T} \tag{25}$$

$$\beta_T = p_{1,T}\,\xi_T + p_{2,T} \tag{26}$$

$$\gamma_T = \xi_T' p_{1,T} \xi_T \tag{27}$$

Integrating the Ricati equations backwards in time we get all the terms needed to compute the parameters $w_{1,t}$, $w_{2,t}$ for the optimal policy that maps states into accelerations. The control policy that maps states into torques follows

$$\boxed{\begin{aligned} &\tau_t(X_t) = k_{1,t}(X_t) + k_{2,t}(X_t)X_t \\[1ex] &k_{1,t}(X_t) = M(\theta_t)w_{1,t} - N(\theta_t, \dot{\theta}_t) \\[1ex] &k_{2,t}(X_t) = M(\theta_t)w_{2,t} \end{aligned}} \tag{28}$$

V. Comparison with Previous Approaches

Computed Torque Control (CTC) in its many forms is arguably the most popular approach for applications of articulated robots [9]. The most general and sophisticated form of CTC includes a closed loop PID controller as well as compensation for Coriolis/Centrifugal forces, gravitational forces, and friction forces. The resulting control law can be expressed as follows [4]

$$\tau_t = M(\theta_t)\Big(\ddot{\xi}_t + K_p e_t + K_v \dot{e}_t + K_i \bar{e}_t\Big) - N(\theta_t, \dot{\theta}_t) \tag{29}$$

where

$$e_t = \xi_t - \theta_t \tag{30}$$

$$\dot{e}_t = \frac{de_t}{dt} \tag{31}$$

$$\bar{e}_t = \int_0^t e_s ds \tag{32}$$

and K_p, K_i, K_d are gain matrices for the proportional, integral an derivative parts of the feedback controller. CTC can be derived from the point of view of feedback linearization [7], [11]. Minimum Angular Acceleration Control (AAC) can also be seen as a form of feedback linearization so in this sense CTC and AAC belong to the same family of control algorithms. Moreover CTC and AAC controllers are affine on the augmented state of joint angles, angular velocities, and integrated joint angles. However standard CTC and AAC differ in important ways. CTC was derived from the point of view of classical feedback control and thus its emphasis is on guaranteeing stability. On the other hand AAC is derived from the point of view of stochastic optimal control. As such AAC controllers are guaranteed to optimize an average performance criterion, rather than guaranteeing stability. In CTC the gains of the closed loop controller need to be tuned using a non-trivial process. In AAC once the design matrices p, and q are specified, the control policy is determined without the need to tune the gains. In truth, it remains to be seen whether tuning p and q is any easier than tuning PID gain matrices. However the most important difference between CTC and AAC is the fact that while CTC is non-anticipatory and utilizes constant gains, AAC is anticipatory and has gain matrices that are a non-linear function of time. As we see in the next section these two properties are responsible for making AAC more energy efficient than CTC and for generating motions that adhere to well known principles of biological motor control.

AAC also belongs to the recent family of control algorithms inspired on the theory of stochastic optimal control. This family includes local iterative solutions to the HJB equation, global approximate solutions using function approximation, and approximate solutions using variations of stochastic gradient descent, as in reinforcement learning approaches [17], [20], [15], [1], [14], [16], [8]. These algorithms can be used to approximate control laws that achieve task goals while minimizing energy consumption. AAC can be seen as one such approximation where we use minimization of angular acceleration as a proxy for minimization of energy consumption.

VI. Computer Simulations

The goal of the simulations was to compare the behavior of CTC, AAC and humans on standard control tasks. In particular we were interested on the trade-offs between energy consumption and task accuracy achieved by CTC and AAC. To this end we simulated a 7 degree of freedom model of the human arm. The first joint (shoulder) had 3 degrees of freedom, the second joint (elbow) had 2 degrees of freedom and the third joint (wrist) had 2 degrees of freedom. The links were simulated as ellipsoids of standard human adult size, with the density of ice. Gravitational forces used the Earth surface standard. The simulator was implemented in Matlab using the Gaussian mechanics approach to articulated bodies [12]. The equations of motion were integrated using a 4th order implicit Runge-Kutta method with a time step of 1 millisecond. Our implementation was validated using the Matlab Robotics Toolbox [5]. The CTC and AAC controllers where also implemented in Matlab. The dispersion matrices that control uncertainty in the system dynamics were set as follows: c_t was diagonal with constant value of 10. The dispersion matrices g_t and h_t were set to zero. The Ricati equations for the optimal value function were integrated using a Backwards Euler approach with a 1 millisecond time step. The human data was obtained by digitizing Figure 3B from [21]. The simulations were run in a MacBookPro laptop. For the tasks presented here finding the optimal AAC controller takes a fraction of a second. The robot arm simulations run comfortably in real time. In all cases we were interested in comparing the energy consumption of different control approaches. Assuming the joints are driven by DC motors, the torque is proportional to the current driving the motor. Moreover the energy consumption is proportional to the square of the current. Thus the integral over time of the squared torques provides a measure of energy consumption (Jules per Ohm) comparable across the different control approaches.

TABLE I

Task Error and Energy Consumption for the Quadratic Torque Control, Quadratic Torque Control plus kinematic Minimum Jerk, and minimum Angular Acceleration Control

	CTC	CTC/MJ	AAC
Error	1481	956	718
Energy	3820	2946	75

We simulated the point-to-point experiment described in Flash & Hogan classic study on minimum jerk kinematic control [6]. The task consisted of moving the end effector (hand) between 6 different points on the horizontal plane. The location of the points on the plane with respect to the arm, the flight times, and approximate levels of variability in human trajectories were obtained by digitizing the figures in [21]. We used this paper rather than the original Flash & Hogan document because it contains data about trajectory variability through time. We compared three different controllers: (1) CTC whose only inputs were the start and end points. (2) CTC controller applied on top of a Minimum Jerk kinematic trajectory planner between the start and end points. We refer to this as CTC/MJ. (3) AAC with zero task cost during flight time between beginning and end points.

For AAC the acceleration cost matrix q_t was the identity matrix throughout the entire flight time. The flight times were set to approximate the times reported in [21]. During flight time the task design matrices p_t were set to zero. The p_t matrices were then made non zero for an additional 0.3 seconds after the flight time ended. In particular the p_t matrices were set to 200000 for the angular terms and 500 for the angular integrals angular velocity terms. These numbers were chosen because they produced trajectories that appeared to qualitatively match the human data.

To set the PID gains for the CTC and for the CT/MJ controllers we performed hierarchical grid search. The goal of this search was to minimize the quadratic task error ϕ. Table VI displays the average error and average energy consumption produced by the three control approaches. Best task performance was obtained by the AAC approach. The average error was 718 and the average energy consumption was 75 Jules/Ohm. Second best was the CTC/MJ approach, with a task error approximately equal to AAC but an energy consumption more than 40 times larger. The worst performer was CTC, with an error about twice as large as AAC and an energy consumption about 50 times larger.

Figure 1 shows sample trajectories and velocity curves for the different control approaches and for the human trajectories in [21]. As is now well known humans exhibit symmetric bell shaped velocity trajectories. Due to the fact that CTC uses constant closed loop gains, it exhibits asymmetric velocity curves with maximum velocity near the start of the movement. CTC/MJ is forced to follow a bell-shaped minimum jerk trajectory known to approximate well human data. However due to the fact that CTC uses constant closed loop gains, there is little room for variability during the flight times. This contrasts with the large variability exhibited by

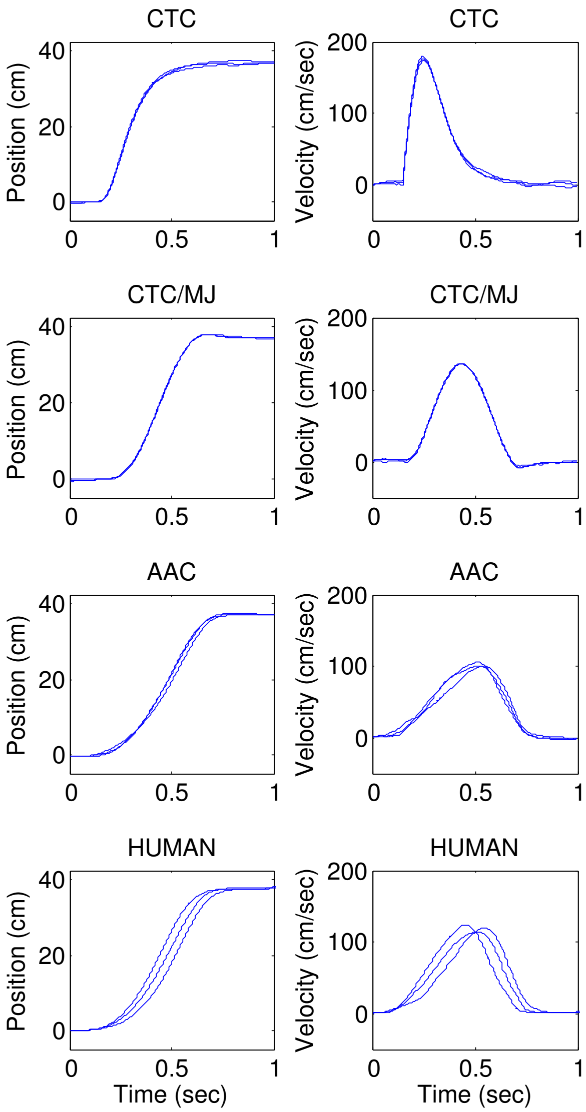

Fig. 1. Left Column: Projection of end effector trajectory on the line between start and end points. Right Column: Velocity profiles. CTC: Computed Torque Control. CTC/MJ: Computer torque control with Minimum Jerk kinematic planner. AAC: Angular Acceleration Control.

humans during flight time. Interestingly AAC chose a bell-shaped curve with levels of variability that were larger at flight time than at the end points.

Figure 2 shows three example trajectories from the 3 controllers and from humans for the T2 to T6 point task in [21], [6]. CTC produces slightly curved trajectories with approximately uniform variability throughout flight time. CTC/MJ produces straight trajectories with uniform variability. Humans and AAC produce curved trajectories with large variability throughout flight time and low variability at the end points, where it matters. This variability profile is a example of the principle of minimum intervention [19]. AAC can produce this variability profile because of the fact that it generates time dependent closed loop gains.

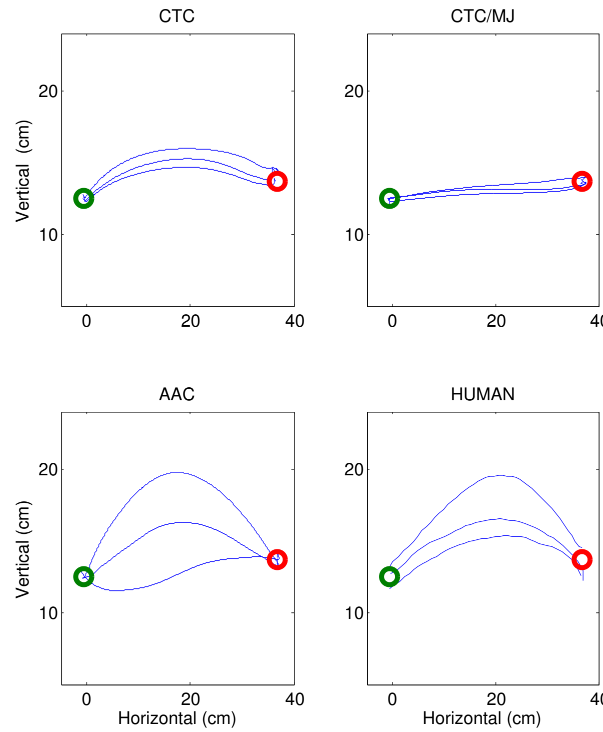

Fig. 2. Sample trajectories for the T2-T6 point to point control problem.

VII. Conclusions

We presented a Minimum Angular Acceleration approach to the closed loop control of articulated body dynamics. The approach obviates the need for kinematic trajectory planners and directly maps joint angle states into torques. The approach is computationally light-weight and results on significant reduction on movement energy consumption when compared to popular approaches such as Computed Torque Control (CTC). Low energy consumption is important in and of itself for mobile robots because it results on longer battery life. More importantly low energy consumption is achieved by using low torques when they do not affect the task at hand. The use of low torques can help improve the safety of robots that operate in human spaces. In addition efficient, low energy trajectories have properties similar to those found in human motion, including anticipatory control, bell-shaped velocity profiles, and larger variability at the mid-points of the trajectories.

The AAC approach proposed here can be seen as an example of feedback linearization of the type used by the popular CTC approach. However contrary to CTC, the AAC approach combines a closed loop controller with time dependent gains and an anticipatory open loop controller. AAC is framed in the theory of stochastic optimal control and thus it also belongs to the recent family of controllers that find approximately optimal trade-offs between task performance and energy consumption [17], [20], [15]. The disadvantage of AAC is that it does not directly optimize energy consumption but angular acceleration. This is particularly important when gravitational forces can be used to save torque. The advantage is that using angular acceleration as a proxy for energy consumption greatly simplifies the control problem: It obviates the need for computationally expensive, iterative algorithms and it still achieves graceful, low energy trajectories that resemble some of the key properties of human motion.

REFERENCES

[1] C.G. Atkeson. Learning arm kinematics and dynamics. *Annual Review of Neuroscience*, 12(1):157–183, 1989.

[2] N.A. Bernstein. *The co-ordination and regulation of movements*. Pergamon Press Ltd., Oxford, 1967.

[3] D. Bertsekas and S. Shreve. *Stochastic Optimal Control*. Athena Scientific, 1996.

[4] Luigi Villani By Bruno Siciliano, Lorenzo Sciavicco. *Robotics: modelling, planning and control*. Springer, 2009.

[5] P.I. Corke. A robotics toolbox for matlab. *IEEE Robotics and Automation Magazine*, 3(1):24–32, 1996.

[6] T. Flash and N. Hogan. The coordination of arm movements: An experimentally confirmed mathematical model. *Journal of Neuroscience*, 5:1688–1703, 1985.

[7] E. G. Gilbert and I. J. Ha. An approach to nonlinear feedback control and its application to robotics. *IEEE Transactions on Systems Man and Cybernetics*, 14(6):879–884, 1984.

[8] Stefan Schaal Jan Peters, Sethu Vijayakumar. Reinforcement learning for humanoid robotics. In *Proceedings of the Third IEEE-RAS International Conference on Humanoid Robots*, 2003.

[9] P.K. Khosla and T. Kanade. Experimental evaluation of nonlinear feedback and feedforward control schemes for manipulators. *International Journal of Robotics Research*, 7(1):18–28, 1988.

[10] Jens Kober and Jan Peters. Reinforcement learning in robotics: a survey. In *Reinforcement Learning*. Springer, 2011.

[11] K. Kreutz. On manipulator control by exact linearization. *IEEE Transactions on Automatic Control*, 34:763–767, 1989.

[12] J. R. Movellan. *Physics for Roboticds and Animation: A Gaussian Approach*. MPLab, UCSD, 2011.

[13] B. Oksendal. *Stochastic Differential Equations: An introduction with applications. Fifth Edition*. Springer Verlag, Berlin, 1998.

[14] Stefan Schaal. Learning from demonstration. In *Advances in Neural Information Processing Systems 9*, pages 1040–1046. MIT Press, 1997.

[15] A. Simpkins and E. Todorov. Practical numerical methods for stochastic optimal control of biological systems in continuous time and space. *Proceedings of the 2nd IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning*, pages 212–218, 2009.

[16] E.A. Theodorou, J. Buchli, and S. Schaal. A generalized path integral control approach to reinforcement learning. *Journal of Machine Learning Research*, 11:3137–3181, 2010.

[17] Evangelos Theodorou, Yuval Tassa, and Emo Todorov. *Stochastic Differential Dynamic Programming*. American Control Conference, 2010.

[18] E. Todorov. Optimal control theory. In Doya K., editor, *Bayesian Brain: Probabilistic Approaches to Neural Coding*. MIT Press, Cambridge, 2006.

[19] E. Todorov and J. I. Jordan. Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, 5:1226–1235, 2002.

[20] E. Todorov and W. Li. A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems. *Proceedings of the American Control Conference*, 2005.

[21] Y. Uno, M. Kawato, and R. Suzuki. Formation and control of optimal trajectory in human multijoint arm movement. *Biological cybernetics*, 61(2):89–101, 1989.