IEEE WCCI 2020
IEEE World Congress on Computational Intelligence
Virtual Conference – July 19-24, 2020

A Bayesian perspective on classical control

Manuel Baltieri

RIKEN CBS (Centre for Brain Science), Saitama, Japan

SPONSORS:

IEEE Advancing Technology for Humanity
IEEE Computational Intelligence Society
IET
The International Neural Network Society (INNS)
EPS Evolutionary Programming Society

Roadmap

- Bayesian readings of optimal control
- Bayesian readings of classical control (work in progress)

  Case 1: PID tuning

  Case 2: PID with multiple degrees of freedom

  Case 3: A general Bayesian framework

Bayesian inference and control

Duality of inference and control, a brief historical account:

- Kalman (Kalman filter - LQR, observability - controllability), 60's
- Johnson (Integral control - polynomial bias estimation), 60-70's
- Fleming and Mitter (Nonlinear filtering into linear control via logarithmic transformation of HJB equation), 70-80's
- Whittle et al. (Risk sensitive control), 80's
- Mitter and Newton (Variational interpretation of optimal control), 00's
- Kappen, Todorov (Applications of logarithmic transformation to control problems, new dualities), 00-10's
- (many more)

Examples in neuroscience and robotics

PNAS PNAS PNAS PNAS

Efficient computation of optimal actions

Emanuel Todorov[1]

Departments of Applied Mathematics and Computer Science & Engineering, University of Washington, Box 352420, Seattle, WA 98195

Edited by James L. McClelland, Stanford University, Stanford, CA, and approved April 28, 2009 (received for review November 16, 2007)

Optimal choice of actions is a fundamental problem relevant to fields as diverse as neuroscience, psychology, economics, computer science, and control engineering. Despite this broad relevance the abstract setting is similar: we have an agent choosing actions over time, an uncertain dynamical system whose state is affected by those actions, and a performance criterion that the agent seeks to optimize. Solving problems of this kind remains hard, in part, because of overly generic formulations. Here, we propose a more structured formulation that greatly simplifies the construction of optimal control laws in both discrete and continuous domains. An exhaustive search over actions is avoided and the problem becomes linear. This yields algorithms that outperform Dynamic Programming and Reinforcement Learning, and thereby solve traditional problems more efficiently. Our framework also enables computations that were not possible before: composing optimal control laws by mixing primitives, applying deterministic methods to stochastic systems, quantifying the benefits of error tolerance, and inferring goals from behavioral data via convex optimization. Development of a general class of easily solvable problems tends to accelerate progress—as linear systems theory has done, for example. Our framework may have similar impact in fields where optimal choice of actions is relevant.

action selection | cost function | linear Bellman equation | stochastic optimal control

Reinforcement Learning (2) methods that are very general but can be inefficient. Indeed, Eq. **1** characterizes $v(x)$ only implicitly, as the solution to an unsolved optimization problem, impeding both analytical and numerical approaches.

Here, we show how the Bellman equation can be greatly simplified. We find an analytical solution for the optimal u given v, and then transform Eq. **1** into a linear equation. Short of solving the entire problem analytically, reducing optimal control to a linear equation is the best one can hope for. This simplification comes at a modest price: although we impose certain structure on the problem formulation, most control problems of practical interest can still be handled. In discrete domains our work has no precursors. In continuous domains there exists related prior work (6–8) that we build on here. Additional results can be found in our recent conference articles (9–11), online preprints (12–14), and supplementary notes [supporting information (SI) *Appendix*].

Results

Reducing Optimal Control to a Linear Problem. We aim to construct a general class of MDPs where the exhaustive search over actions is replaced with an analytical solution. Discrete optimization problems rarely have analytical solutions, thus our agenda calls for continuous actions. This may seem counterintuitive if one thinks of actions as symbols ("go left," "go right"). However, what gives meaning to such symbols are the underlying transition probabilities—which are continuous. The latter observation is key

Journal of Machine Learning Research 11 (2010) 3137-3181 Submitted 1/10; Revised 7/10; Published 11/10

A Generalized Path Integral Control Approach to Reinforcement Learning

Evangelos A.Theodorou ETHEODOR@USC.EDU
Jonas Buchli JONAS@BUCHLI.ORG
Stefan Schaal* SSCHAAL@USC.EDU
Department of Computer Science
University of Southern California
Los Angeles, CA 90089-2905, USA

Editor: Daniel Lee

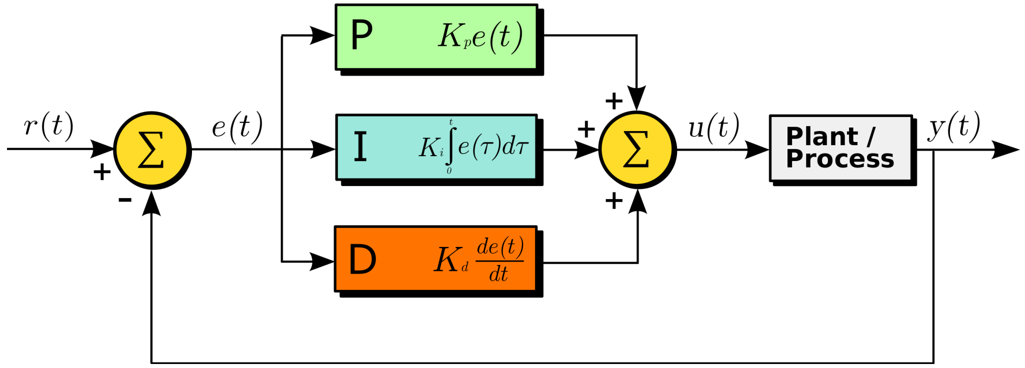
Abstract

With the goal to generate more scalable algorithms with higher efficiency and fewer open parameters, reinforcement learning (RL) has recently moved towards combining classical techniques from optimal control and dynamic programming with modern learning techniques from statistical estimation theory. In this vein, this paper suggests to use the framework of stochastic optimal control with path integrals to derive a novel approach to RL with parameterized policies. While solidly grounded in value function estimation and optimal control based on the stochastic Hamilton-Jacobi-Bellman (HJB) equations, policy improvements can be transformed into an approximation problem of a path integral which has no open algorithmic parameters other than the exploration noise. The resulting algorithm can be conceived of as model-based, semi-model-based, or even model free, depending on how the learning problem is structured. The update equations have no danger of numerical instabilities as neither matrix inversions nor gradient learning rates are required. Our new algorithm demonstrates interesting similarities with previous RL research in the framework of probability matching and provides intuition why the slightly heuristically motivated probability matching approach can actually perform well. Empirical evaluations demonstrate significant performance improvements over gradient-based policy learning and scalability to high-dimensional

What can classical control gain from a Bayesian perspective?

Case 1 - Tuning rules

- In Baltieri & Buckley (2019), we re-derived PID control equations as a problem of approximate variational inference (cf. Johnson and his derivation of integral control as bias estimation in observer models)
- PID gains as precisions (i.e., inverse variances) on observations and their higher embedding orders
- A new algorithm to optimise PID gains based on second order optimisation of a variational free energy (i.e., first order minimisation of a variational *action*, see Friston (2008))

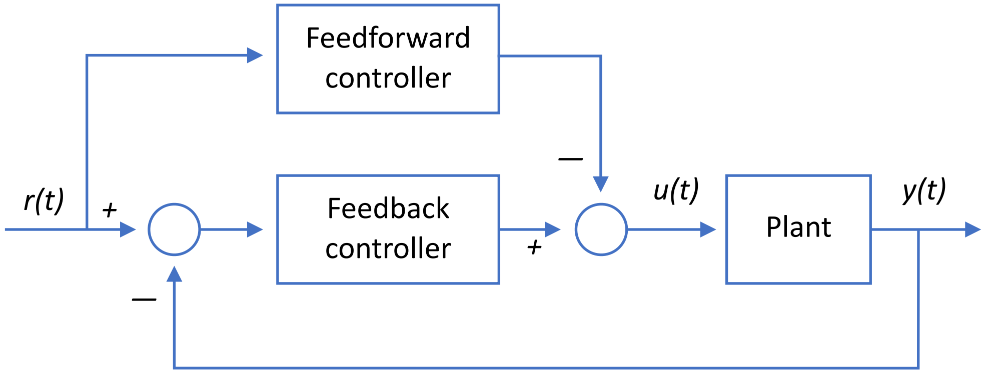


(Image courtesy of Wikimedia Commons)

$$u(t) = k_p e(t) + k_i \int_0^t e(\tau)d\tau + k_d \frac{de(t)}{dt}$$

Case 2 - Theoretical background

- A classical problem in controller design: two degrees of freedom to differentiate set-point changes from load disturbances
- The standard solution: each degree of freedom requires an independently adjusted closed-loop transfer function
- Usually this is limited to two degrees of freedom (2DOF) with a design including a feedforward and feedback component
- Our reading: feedforward and feedback components correspond to prior (before observations) and likelihood (after observations) in a variational scheme. More next.



2DOF PID control

Case 3 - Performance-robustness trade-off

Following Åström & Hägglund (2000).

Performance criteria:

- Load disturbance response
- Set-point response
- Measurement noise response

Robustness criteria

- Robustness to model uncertainty

A proposal: state-space formulation

State space model

Dynamics	$\tilde{\mu}'_x = f(\tilde{\mu}_x, \tilde{\mu}_v) + \tilde{w}$	System fluctuations	$\tilde{w} \sim N(0, \pi_{\tilde{w}})$
Observations	$\tilde{y} = g(\tilde{\mu}_x, \tilde{\mu}_v) + \tilde{z}$	Observation noise	$\tilde{z} \sim N(0, \pi_{\tilde{z}})$

(Stochastic volatility + time-scale separation assumptions)

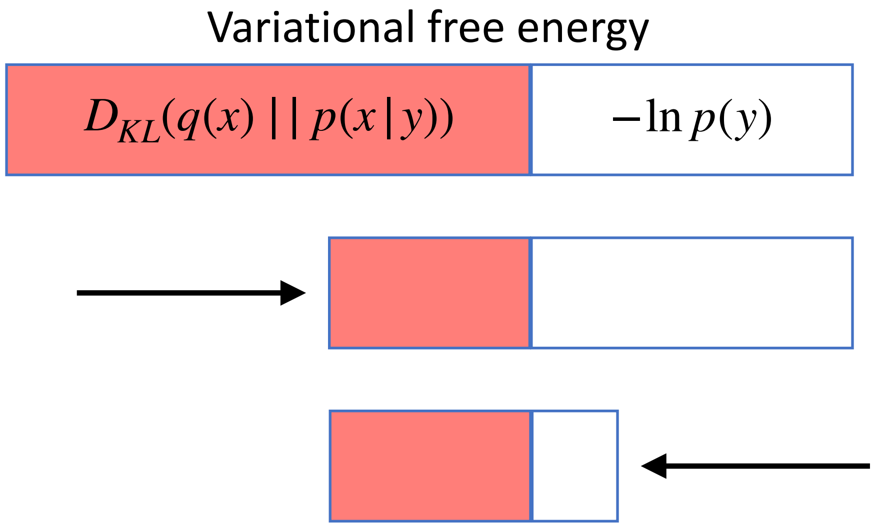
System precision law	$\pi_{\tilde{w}} = k(\eta_{\tilde{w}}) + \tilde{r}_w$	Uncertainty on system precision	$\tilde{r}_w \sim N(0, p_{\tilde{w}})$
Observation precision law	$\pi_{\tilde{z}} = h(\eta_{\tilde{z}}) + \tilde{r}_z$	Uncertainty on observation precision	$\tilde{r}_z \sim N(0, p_{\tilde{z}})$

A variational formulation of classical controllers - Recipe

- Describe a standard problem of inference given some goals, i.e., find posterior beliefs given a likelihood function and some prior encoding reference target

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

- Write down variational bound (variational free energy or ELBO) to approximate the posterior
- Minimise variational free energy to approximate posterior (standard variational Bayes)
- At the same time, update observations by acting/ controlling them, i.e., change surprisal while minimising bound



SPONSORS:
IEEE WCCI 2020
IEEE World Congress on Computational Intelligence
Virtual Conference – July 19-24, 2020
IEEE Advancing Technology for Humanity
IEEE Computational Intelligence Society
IET
The International Neural Network Society (INNS)
EPS Evolutionary Programming Society

A schematic recipe

Generative model
-
A probabilistic rewriting of the previous SSM

Variational free energy for approximate Bayesian inference
(with variational Gaussian and Laplace assumptions)

Gradient descent on free energy given a generative model

Recognition dynamics
-
Controller

$\dot{u}(t)$

$y(t)$

Generative process
-
Plant

SPONSORS:
IEEE WCCI 2020
IEEE World Congress on Computational Intelligence
Virtual Conference – July 19-24, 2020
IEEE Advancing Technology for Humanity
IEEE Computational Intelligence Society
IET
The International Neural Network Society (INNS)
EPS Evolutionary Programming Society

The variational free energy

$$F \approx \frac{1}{2}\Bigg[\underbrace{\mu_{\pi_{\tilde{z}}}\Big(\tilde{y} - g(\tilde{\mu}_x, \tilde{\mu}_v)\Big)^2}_{\text{load disturbances}} + \underbrace{\mu_{\pi_{\tilde{w}}}\Big(\tilde{\mu}'_x - f(\tilde{\mu}_x, \tilde{\mu}_v)\Big)^2}_{\text{set-point changes}} + \underbrace{\mu_{p_{\tilde{z}}}\Big(\mu_{\pi_{\tilde{z}}} - h(\eta_{\pi_{\tilde{z}}})\Big)^2}_{\text{meas. noise response}} + \underbrace{\mu_{p_{\tilde{w}}}\Big(\mu_{\pi_{\tilde{w}}} - k(\eta_{\pi_{\tilde{w}}})\Big)^2}_{\text{model uncertainty}}\Bigg]$$

- The first two terms describe a standard 2DOF PID controller when $f(\tilde{\mu}_x, \tilde{\mu}_v) = r(t)$, see Baltieri & Buckley, 2019
- The third term provides a prior on stochastic volatility of measurement noise, e.g., when a sensor breaks, this can be used to update over time the best estimate of current measurement noise, (cf. empirical and hierarchical Bayes)
- The fourth term provides a prior for stochastic volatility of system noise, e.g., model uncertainty can change over time (decreasing when new info. becomes available) and dedicated priors may encode info. for a *class* of systems, (cf. empirical and hierarchical Bayes)

Minimisation scheme

$$F \approx \frac{1}{2}\Big[\underbrace{\mu_{\pi_{\tilde{z}}}\big(\tilde{y} - g(\tilde{\mu}_x, \tilde{\mu}_v)\big)^2}_{\text{load disturbances}} + \underbrace{\mu_{\pi_{\tilde{w}}}\big(\tilde{\mu}'_x - f(\tilde{\mu}_x, \tilde{\mu}_v)\big)^2}_{\text{set-point changes}} + \underbrace{\mu_{p_{\tilde{z}}}\big(\mu_{\pi_{\tilde{z}}} - h(\eta_{\pi_{\tilde{z}}})\big)^2}_{\text{meas. noise response}} + \underbrace{\mu_{p_{\tilde{w}}}\big(\mu_{\pi_{\tilde{w}}} - k(\eta_{\pi_{\tilde{w}}})\big)^2}_{\text{model uncertainty}}\Big]$$

Minimise variational free energy

$$\dot{\tilde{\mu}}_x = \nabla_{\mu_x} F$$

$$\dot{u}(t) = \nabla_u F = \frac{\partial F}{\partial y}\frac{\partial y}{\partial u}$$

$$\dot{\mu}_{\pi_{\tilde{z}}} = \nabla_{\mu_{\pi_{\tilde{z}}}} F$$

$$\dot{\mu}_{\pi_{\tilde{w}}} = \nabla_{\mu_{\pi_{\tilde{w}}}} F$$

- Minimise variational bound balancing 2DOF (first two weighted prediction errors)
- Build PID controller to output u combining 2DOF
- Optimise stochastic volatility of measurement noise
- Optimise stochastic volatility of system noise (model uncertainty)

Intuitive use of hierarchical and empirical Bayes

Hierarchical Bayes: turn all parameters and hyperparameters into random variables, use their known uncertainty to our advantage

For example, if there is prior knowledge that a system to be controlled has a certain fault tolerance, embed this information in the controller's model to allow moving to a new operating regime (e.g., a wheeled robot losing the ability to utilise one wheel)

Empirical Bayes: use data/observations to estimate priors

For example, given a class of systems to control (e.g., all machines with a certain engine), determine the most likely **class** priors to be used as initial conditions for control, states and/or parameters

Summary

- A probabilistic framework to describe classical controllers as variational inference
- More theoretical background (2DOF controllers), tuning rules (gradient descent on integral of variational free energy, or local second order scheme)
- Generalisations to different problems: the performance-robustness trade-off of PID control as a hierarchical Bayesian scheme
- Implementations hopefully soon
- (Collaborations?)

Acknowledgements:

Japanese Society for the Promotion of Science (JSPS) - Postdoctoral fellowships program.

This project was partially supported by a JSPS Grant-in-Aid for Scientific Research (No. 19F19809).

Thank you

Contact:

Email: manuel.baltieri [at] riken.jp

Twitter: @manuelbaltieri