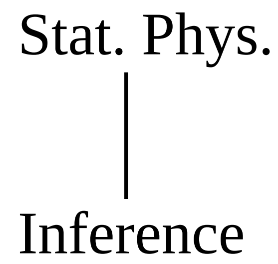
The statistical physics of optimal control theory

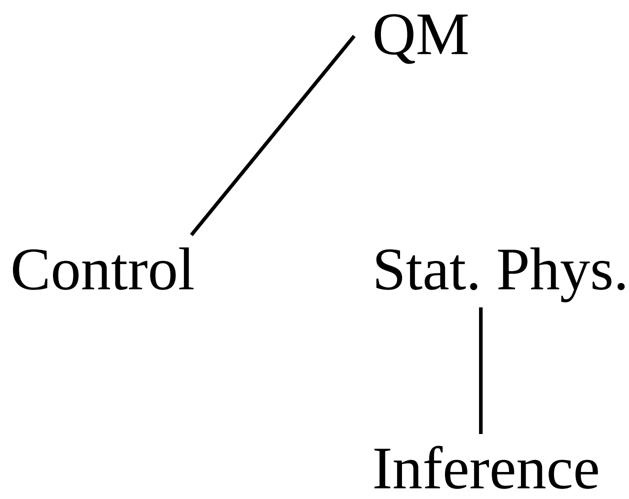
Bert Kappen
Radboud University Nijmegen
University College London

September 2012

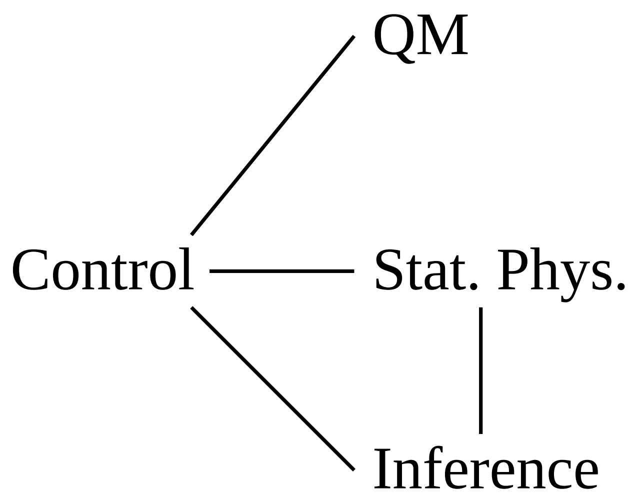
The statistical physics of control and inference



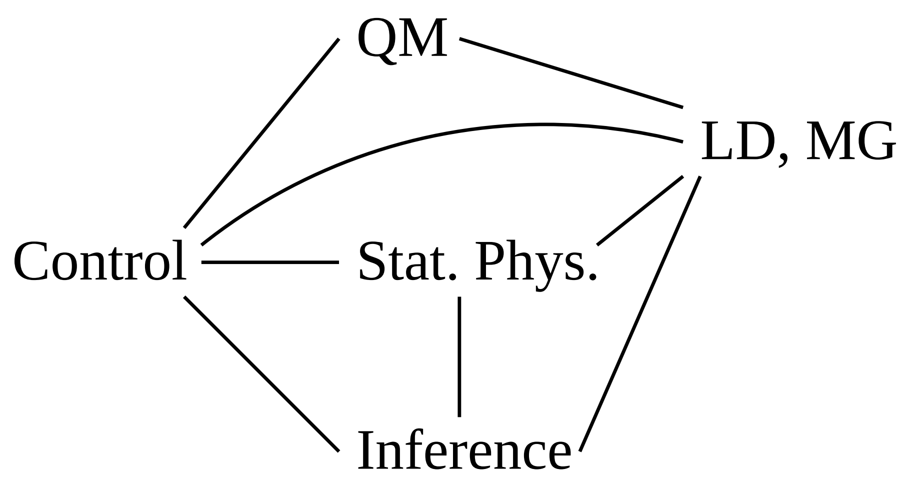
The statistical physics of control and inference



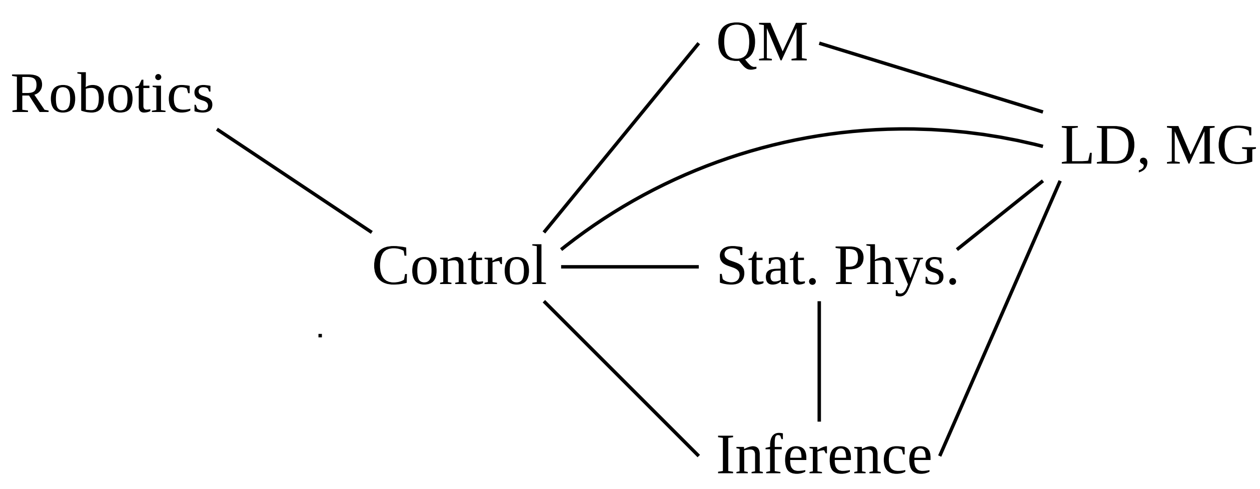
The statistical physics of control and inference



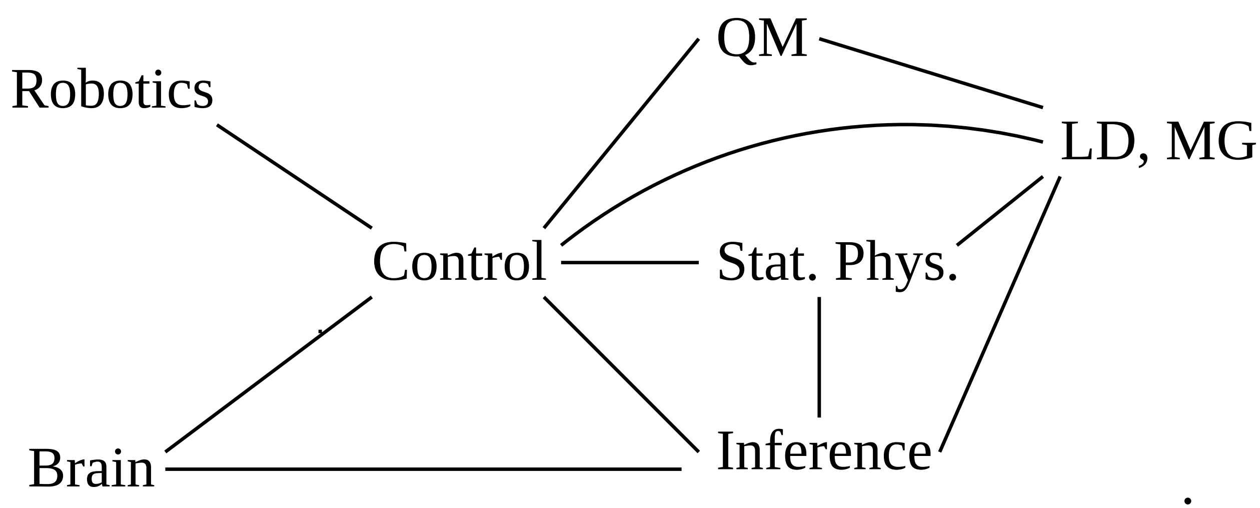
The statistical physics of control and inference



The statistical physics of control and inference



The statistical physics of control and inference



Outline

- Link between control theory and quantum mechanics
  - Mandelung '27, Nelson '67, Guerra '81 and others

Outline

- Link between control theory and quantum mechanics
  - Mandelung '27, Nelson '67, Guerra '81 and others
- Link between control theory, inference and statistical physics
  - Hopf '50, Fleming Mitter '82, Kappen '05

Outline

- Link between control theory and quantum mechanics
  - Mandelung '27, Nelson '67, Guerra '81 and others
- Link between control theory, inference and statistical physics
  - Hopf '50, Fleming Mitter '82, Kappen '05
- Link between control theory and large deviations
  - Schrödinger '32

Control and Quantum Mechanics

Schrödinger equation

$$\hbar i \partial_t \Psi(x,t) = V(x) - \frac{\hbar^2}{2}\nabla^2 \Psi(x,t)$$

Write

$$\Psi = \sqrt{\rho} \exp\left(i\frac{S}{\hbar}\right)$$

then

$$\begin{aligned} -\partial_t \rho &= \nabla(\rho \nabla S) \\ -\partial_t S &= \frac{1}{2}(\nabla S)^2 - \frac{1}{2}\hbar^2 \frac{\nabla^2 \sqrt{\rho}}{\sqrt{\rho}} + V \end{aligned}$$

Pontryagin minimum principle

Consider the (deterministic) control problem

$$\dot{x} = f(x,u) \qquad C(u|x_0) = x(T)\phi + \int_0^T dt R(x(t), u(t))$$

We find the optimal control by constrained optimization.

Define the Lagrangian with λ the co-state

$$L(x,u,\lambda) = \lambda f(x,u) - R(x,u) \qquad \mathcal{C} = \int_0^T dt L(x(t), u(t), \lambda(t))$$

$\delta\mathcal{C} = 0$ implies

$$\begin{aligned} H(x,\lambda) &= \mathrm{argmax}_u L(x,u,\lambda) \\ \dot{x} &= \partial_\lambda H(x,\lambda) \qquad x(0) = x_0 \\ \dot{\lambda} &= -\partial_x H(x,\lambda) \qquad \lambda(T) = -\phi \end{aligned}$$

Control and Quantum Mechanics

Identify $\rho(x)$ with the state and $S(x)$ with the co-state and define

$$
\begin{aligned}
H(\rho, S) &= \int dx \rho(x,t) \left(\frac{1}{2}(\nabla S(x,t))^2 + V(x) - \frac{1}{8}\hbar^2 \nabla^2 \log \rho(x,t) \right) \\
\partial_t \rho &= \partial_S H(\rho, S) = -\nabla(\rho \nabla S) \\
\partial_t S &= -\partial_\rho H(\rho, S) = -\frac{1}{2}(\nabla S)^2 - V(x) + \frac{1}{2}\hbar^2 \frac{\nabla^2 \sqrt{\rho}}{\sqrt{\rho}}
\end{aligned}
$$

Control and Quantum Mechanics

Identify $\rho(x)$ with the state and $S(x)$ with the co-state and define

$$
\begin{aligned}
H(\rho, S) &= \int dx \rho(x,t) \left(\frac{1}{2}(\nabla S(x,t))^2 + V(x) - \frac{1}{8}\hbar^2 \nabla^2 \log \rho(x,t) \right) \\
\partial_t \rho &= \partial_S H(\rho, S) = -\nabla(\rho \nabla S) \\
\partial_t S &= -\partial_\rho H(\rho, S) = -\frac{1}{2}(\nabla S)^2 - V(x) + \frac{1}{2}\hbar^2 \frac{\nabla^2 \sqrt{\rho}}{\sqrt{\rho}}
\end{aligned}
$$

QM satisfies Hamilton dynamics with 'position' ρ and 'momentum' S.

QM as control of distributions

Identify $\rho(x)$ with the state and $S(x)$ with the co-state and define

$$
\begin{aligned}
H(\rho, S) &= \int dx \rho(x,t) \left(\frac{1}{2} (\nabla S(x,t))^2 + V(x) - \frac{1}{8} \hbar^2 \nabla^2 \log \rho(x,t) \right) \\
\partial_t \rho &= \partial_S H(\rho, S) \qquad \rho(x,0) = \rho_0(x) \\
\partial_t S &= -\partial_\rho H(\rho, S) \qquad S(x,T) = -\phi(x)
\end{aligned}
$$

QM satisfies PMP equations of optimal control

QM as control of distributions

Identify $\rho(x)$ with the state and $S(x)$ with the co-state and define

$$
\begin{aligned}
H(\rho, S) &= \int dx \rho(x,t) \left(\frac{1}{2}(\nabla S(x,t))^2 + V(x) - \frac{1}{8}\hbar^2 \nabla^2 \log \rho(x,t)\right) \\
\partial_t \rho &= \partial_S H(\rho, S) \qquad \rho(x,0) = \rho_0(x) \\
\partial_t S &= -\partial_\rho H(\rho, S) \qquad S(x,T) = -\phi(x)
\end{aligned}
$$

QM satisfies PMP equations of optimal control

$$
\begin{aligned}
f(\rho, v) &= -\nabla(v(x,t)\rho(x,t)) \\
C(v|\rho_0) &= \int dx \rho(x,T)\phi(x) \\
&+ \int_0^T dt \int dx \rho(x,t) \left(\frac{1}{2}v(x,t)^2 - V(x) - \frac{1}{8}\nu^2 (\nabla \log \rho(x,t))^2\right)
\end{aligned}
$$

Stochastic interpretation

Consider the dynamical system

$$dx = v_+ dt + d\xi \qquad \left\langle d\xi^2 \right\rangle = \hbar dt$$

The mean and osmotic velocities are

$$v = \frac{1}{2}(v_+ + v_-) \qquad u = \frac{1}{2}(v_+ - v_-) = \frac{1}{2}\hbar \nabla \log \rho$$

The control problem in v_+ becomes

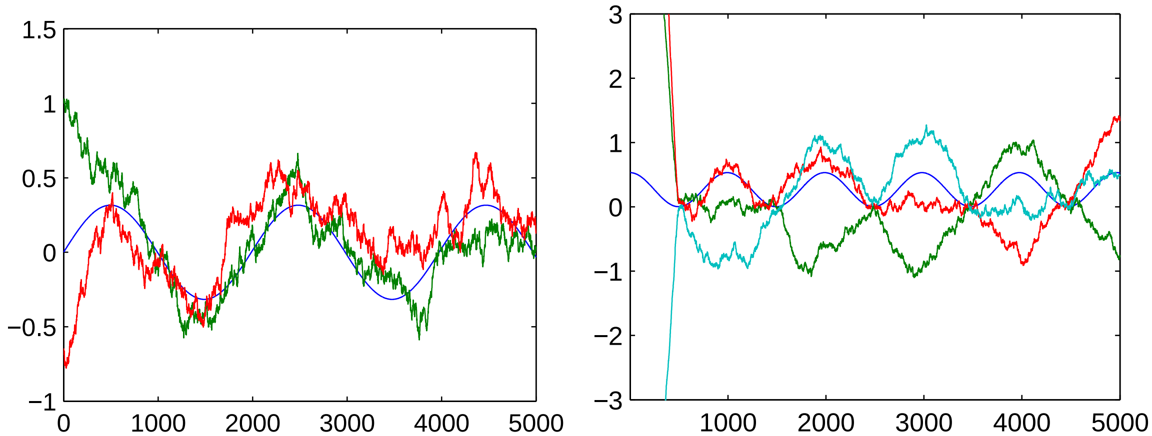
$$\begin{aligned}
\frac{\partial \rho}{\partial t} &= -\nabla(v_+ \rho) + \frac{1}{2}\hbar \nabla^2 \rho = -\nabla(v\rho) \\
C(v|\rho_0) &= \int dx \rho(x,T)\phi(x) \\
&+ \int_0^T dt \int dx \rho(x,t) \left(\frac{1}{2} v_+(x,t) v_-(x,t) - V(x) \right)
\end{aligned}$$

Control and Quantum Mechanics

- Nelson: "Stochastic mechanics is QM made difficult".
  - Computing the drift $v = \nabla S$ requires solving the Schrödinger equation

$$dx = v_+ dt + d\xi \qquad \frac{\partial \rho}{\partial t} = -\nabla(v_+\rho) + \frac{1}{2}\hbar\nabla^2\rho$$

$$C = \left\langle \phi + \int_0^T dt \left(\frac{1}{2} v_+ v_- - V \right) \right\rangle$$



Control and Quantum Mechanics

- Nelson: "Stochastic mechanics is QM made difficult".
  - Computing the drift $v = \nabla S$ requires solving the Schrödinger equation
- What about: "Quantum mechanics is stochastic optimal control made easy?"
  - The control formulation has mixed boundary conditions. Not easily related to boundary conditions on Ψ.
  - The control cost $v_+ v_-$ is a non-linear function of the density

Outline

- Link between control theory and quantum mechanics
  - Mandelung '27, Nelson '67, Guerra '81 and others
- Link between control theory, inference and statistical physics
  - Hopf '50, Fleming Mitter '82, Kappen '05
- Link between control theory and large deviations
  - Schrödinger '32

Link between inference and statistical physics

$p(x_{1:n}) = \pi(x_{1:n})/Z$ is a probability distribution, compute

$$p(x_1) = \sum_{x_{2:n}} p(x_{1:n})$$

Link between inference and statistical physics

$p(x_{1:n}) = \pi(x_{1:n})/Z$ is a probability distribution, compute

$$p(x_1) = \sum_{x_{2:n}} p(x_{1:n})$$

Define free energy

$$F(q) = \sum_{x_{1:n}} q(x_{1:n}) \log \frac{q(x_{1:n})}{\pi(x_{1:n})}$$

F is minimized by $q = p$.

Link between inference and statistical physics

$p(x_{1:n}) = \pi(x_{1:n})/Z$ is a probability distribution, compute

$$p(x_1) = \sum_{x_{2:n}} p(x_1, x_{2:n})$$

Define free energy

$$F(q) = \sum_{x_{1:n}} q(x_{1:n}) \log \frac{q(x_{1:n})}{\pi(x_{1:n})}$$

F is minimized by $q = p$.

Restrict minimization to simple distributions $q(x_{1:n}) = q_1(x_1) \ldots q_n(x_n)$ and minimize

$$p(x_1) \approx q_1(x_1)$$

Link between inference and statistical physics

Efficient inference methods in machine learning:

- Variational/mean field methods, TAP
- Belief propagation, EP, Cluster Variation Method, Survey propagation
- convex relaxations
- Monte Carlo Sampling

Link between control and inference

General idea:

- Express the control problem as an inference problem
- Treat the inference computation as a statistical physics problem

In particular:

- Consider a class of control problems for which the Bellman equation can be transformed in a linear pde (using a log transform)
- 'Solve' as a Feynman-Kac path integral

Path integral control theory

$$
\begin{aligned}
dx_i &= f_i(x,t)dt + \sum_a g_{ia}(x,t)(u_a dt + d\xi_a) \\
C &= \left\langle \phi(x(T)) + \int_t^T ds V(x,t) + \frac{1}{2}\sum_{ab} R_{ab} u_a u_b \right\rangle
\end{aligned}
$$

with $\langle d\xi_a d\xi_b \rangle = \nu_{ab} dt$ and $R = \lambda \nu^{-1}, \lambda > 0$.

The HJB equation becomes

$$
-\partial_t J = \min_u \left(\frac{1}{2} u^T R u + V + (f + gu)^T (\nabla J) + \frac{1}{2} \mathrm{Tr}\left(g \nu g^T \nabla^2 J \right) \right)
$$

with boundary condition $J(x,T) = \phi(x)$.

Path integral control theory

Minimization the HJB equation wrt u yields:

$$
\begin{aligned}
u &= -R^{-1} g^T \nabla J \\
-\partial_t J &= -\frac{1}{2} (\nabla J)^T g R^{-1} g^T (\nabla J) + V + f^T \nabla J + \frac{1}{2} \mathrm{Tr}\left(g \nu g^T \nabla^2 J \right)
\end{aligned}
$$

Define $\psi(x,t)$ through $J(x,t) = -\lambda \log \psi(x,t)$, the HJB becomes *linear* in ψ

$$
\partial_t \psi = \left(\frac{V}{\lambda} - f^T \nabla - \frac{1}{2} \mathrm{Tr}\left(g \nu g^T \nabla^2 \right) \right) \psi
$$

with end condition $\psi(x,T) = \exp(-\phi(x)/\lambda)$.

Feynman-Kac formula

Denote $Q(\tau|x,s)$ the distribution over uncontrolled trajectories that start at x,t:

$$dx = f(x,t)dt + g(x,t)d\xi$$

with τ a trajectory $x(t \to T)$. Then

$$\begin{aligned}\psi(x,t) &= \int dQ(\tau|x,t)\exp\left(-\frac{S(\tau)}{\lambda}\right) \\ S(\tau) &= \phi(x(T)) + \int_t^T ds V(x(s),s)\end{aligned}$$

ψ can be computed by forward sampling the uncontrolled process.

Posterior distribution over optimal trajectories

$\psi(x,t)$ can be interpreted as a partition sum for the distribution over paths under optimal control:

$$P(\tau|x,t) = \frac{1}{\psi(x,t)} Q(\tau|x,t) \exp\left(-\frac{S(\tau)}{\lambda}\right)$$

The optimal cost-to-go is a free energy:

$$J(x,t) = -\lambda \log \int dQ(\tau|x,t) \exp\left(-\frac{1}{\lambda} S(\tau))\right)$$

The optimal control is an expectation wrt P:

$$u(x,t)dt = -R^{-1} g^T(x,t) \nabla J(x,t) dt = \int dP(\tau) d\xi(\tau) = \langle d\xi \rangle_P$$

Recap

Control problem:

$$dx = f dt + g(udt + d\xi) \qquad C = \left\langle \phi + \int_t^T V + \frac{1}{2} u^T R u \right\rangle \qquad R = \lambda \nu^{-1}$$

HJB is linear:

$$\partial_t \psi = H\psi \qquad J = -\lambda \log \psi$$

Solution is given by Feynman-Kac formula: $\psi = \int dQ(\tau) \exp\left(-\frac{S(\tau)}{\lambda}\right)$.
Q distribution over uncontrolled dynamics ($u = 0$).

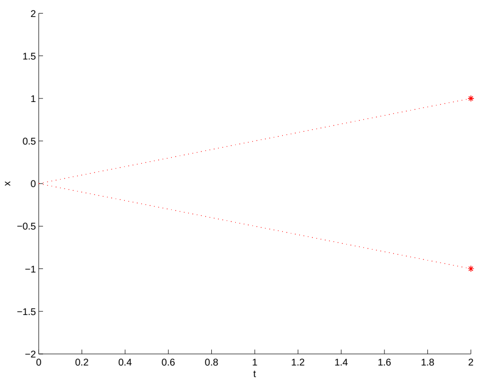
Posterior distribution over optimal controlled trajectories: $P(\tau) = \frac{1}{\psi} Q(\tau) \exp\left(-\frac{S(\tau)}{\lambda}\right)$.

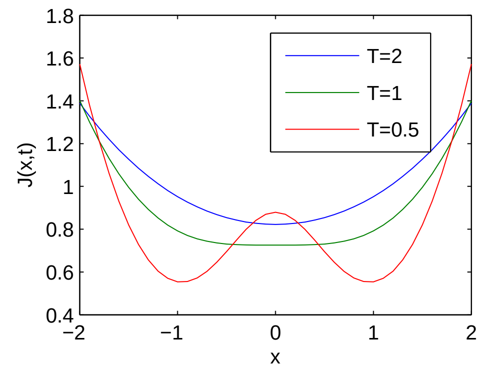
Optimal control is expectation value: $udt = \langle d\xi \rangle_P$.

Delayed choice

$$dx = udt + d\xi \qquad \langle \xi^2 \rangle = \nu dt$$

$V = 0$, path cost is $\frac{1}{2}u^2$, $\phi(x = \pm 1) = 0$ and $\phi(x) = \infty$, else.





"When the future is uncertain, delay your decisions."

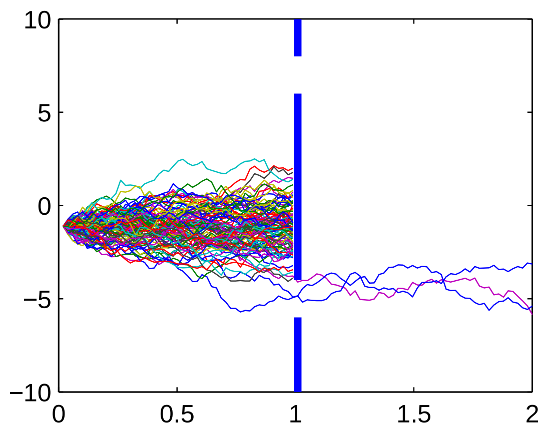
Estimating optimal control by sampling

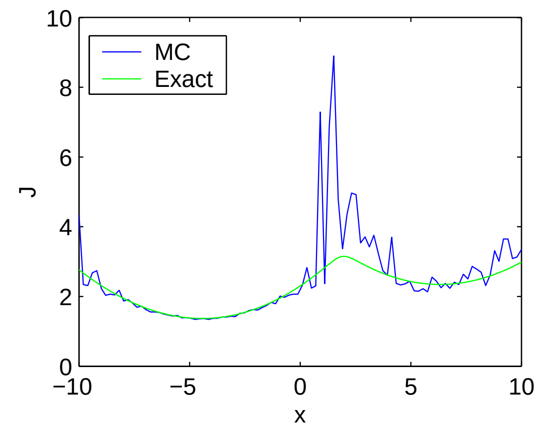
We generate N trajectories $x^\mu_{t:T}$ from $Q(\tau)$ starting at x, t with initial noise $d\xi^\mu$.

The optimal control at x, t is given by (optimal cost-to-go similar)

$$udt = \int dP(\tau)d\xi(\tau) \approx \frac{\sum_\mu d\xi^\mu \exp(-S^\mu/\lambda)}{\sum_\mu \exp(-S^\mu/\lambda)}$$

$$S^\mu = \sum_{s=t}^{T} V(x^\mu_s, s)dt + \phi(x^\mu_T)$$





Unbiased, but inefficient.

Importance sampling

Efficiency may be improved by sampling with $u \neq 0$.

$$\psi(x,t) = \int dQ(\tau) \exp(-S(\tau)/\lambda) = \int dQ'(\tau) \frac{dQ(\tau)}{dQ'(\tau)} \exp(-S(\tau)/\lambda)$$

with $Q'(\tau|x,t)$ from the stochastic process

$$dx = f(x,t)dt + g(x,t)(\hat{u}(x,t)dt + d\xi)$$

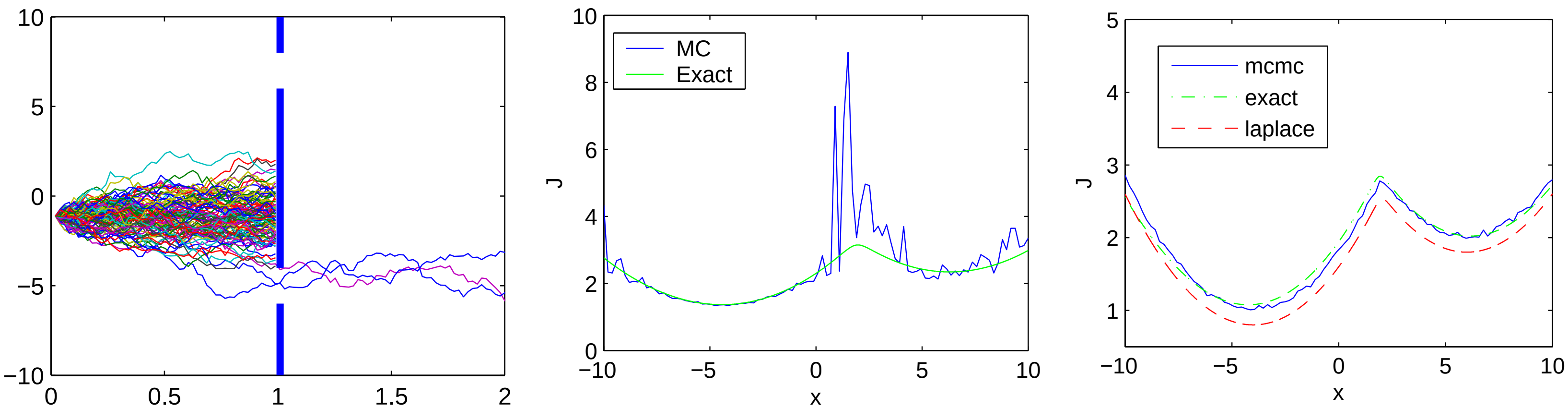
Importance sampling

Efficiency may be improved by sampling with $u \neq 0$.

$$\psi(x,t) = \int dQ(\tau) \exp(-S(\tau)/\lambda) = \int dQ'(\tau) \frac{dQ(\tau)}{dQ'(\tau)} \exp(-S(\tau)/\lambda)$$

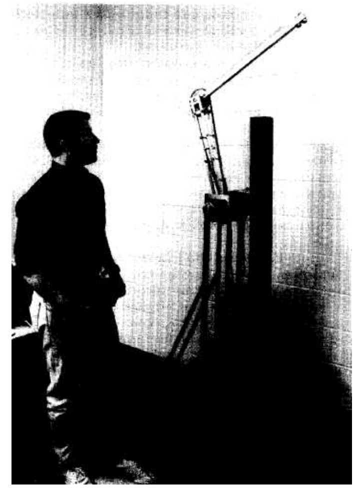
with $Q'(\tau|x,t)$ from the stochastic process

$$dx \quad = \quad f(x,t)dt + g(x,t)(\hat{u}(x,t)dt + d\xi)$$



How to control a device?

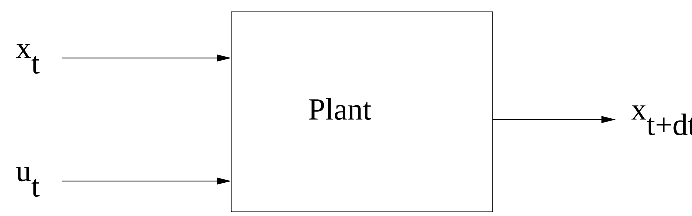
Plant and costs are unknown



$$dx_i = f_i(x,t)dt + \sum_a g_{ia}(x,t)u_a dt$$

$$C = \int_0^T \frac{1}{2}u^T R u + V(x,t)$$

Motor babbling: Generate sequence of states with random control

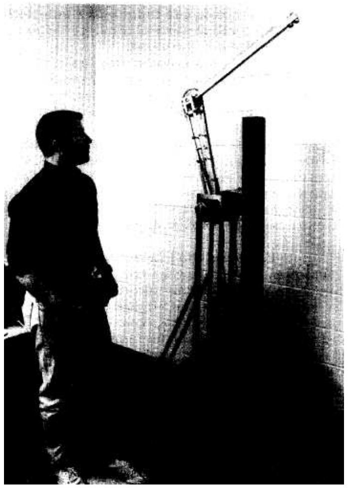


$$u^\mu_{0:T}, x^\mu_{0:T}, \qquad \mu = 1, \ldots, N$$

$$u_a dt = d\xi_a, \nu = \lambda R^{-1}.$$

How to control a device?

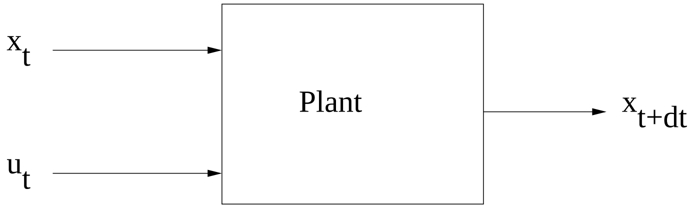
Plant and costs are unknown



$$dx_i = f_i(x,t)dt + \sum_a g_{ia}(x,t)u_a dt$$

$$C = \int_0^T \frac{1}{2}u^T R u + V(x,t)$$

Motor babbling: Generate sequence of states with random control



$$u^{\mu}_{0:T}, x^{\mu}_{0:T}, \qquad \mu = 1,\ldots,N$$

$$u_a dt = d\xi_a, \nu = \lambda R^{-1}.$$

Babbling asymptotically computes the optimal stochastic control

Acrobot

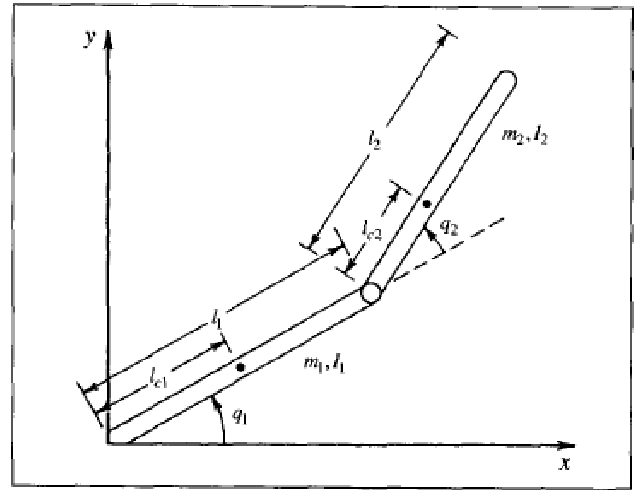


Fig. 1. The Acrobot.

$$d_{11}\ddot{q}_1 + d_{12}\ddot{q}_2 + h_1 + \phi_1 = 0 \qquad (1)$$

$$d_{21}\ddot{q}_1 + d_{22}\ddot{q}_2 + h_2 + \phi_2 = \tau, \qquad (2)$$

where

$$d_{11} = m_1 l_{c1}^2 + m_2(l_1^2 + l_{c2}^2 + 2l_1 l_{c2}\cos(q_2)) + I_1 + I_2$$

$$d_{22} = m_2 l_{c2}^2 + I_2$$

$$d_{12} = m_2(l_{c2}^2 + l_1 l_{c2}\cos(q_2)) + I_2$$

$$d_{21} = m_2(l_{c2}^2 + l_1 l_{c2}\cos(q_2)) + I_2$$

$$h_1 = -m_2 l_1 l_{c2}\sin(q_2)\dot{q}_2^2 - 2m_2 l_1 l_{c2}\sin(q_2)\dot{q}_2\dot{q}_1$$

$$h_2 = m_2 l_1 l_{c2}\sin(q_2)\dot{q}_1^2$$

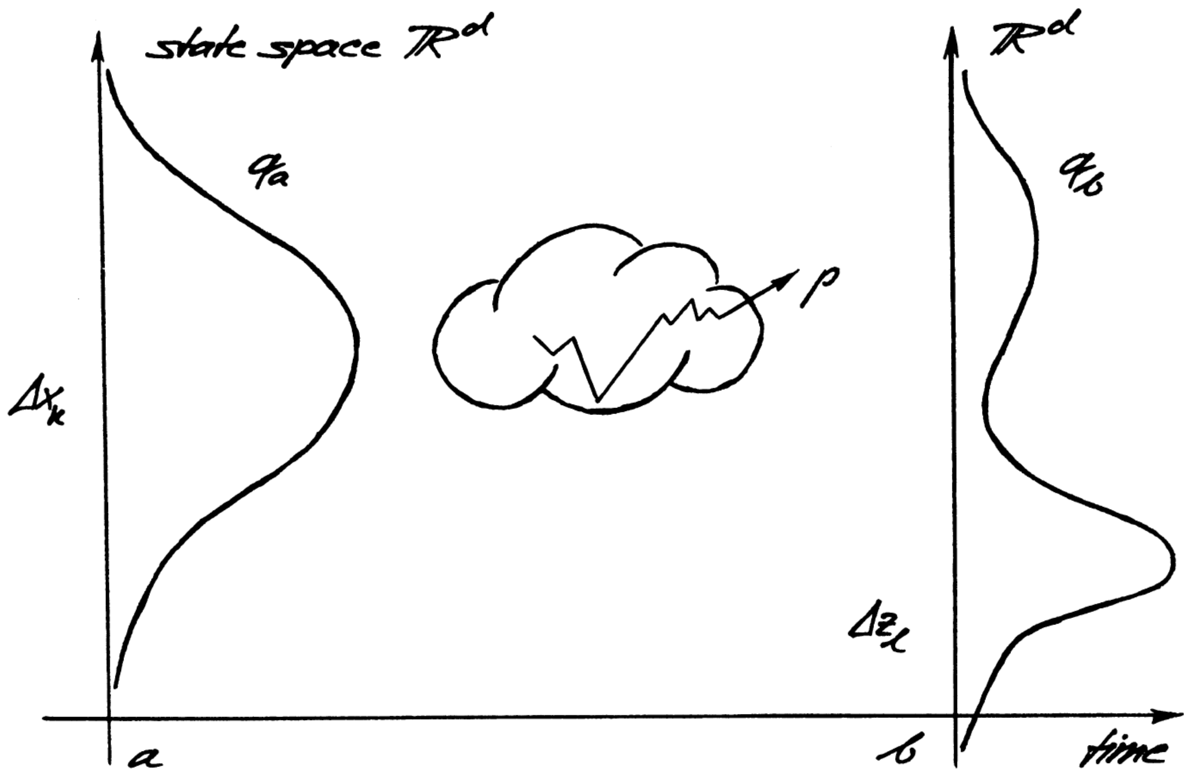
$$\phi_1 = (m_1 l_{c1} + m_2 l_1)g\cos(q_1) + m_2 l_{c2} g\cos(q_1 + q_2)$$

$$\phi_2 = m_2 l_{c2} g\cos(q_1 + q_2).$$

Outline

- Link between control theory and quantum mechanics
  - Mandelung '27, Nelson '67, Guerra '81 and others
- Link between control theory, inference and statistical physics
  - Hopf '50, Fleming Mitter '82, Kappen '05
- Link between control theory and large deviations
  - Schrödinger '32

Link between control theory and large deviations



The Schrödinger bridge problem: given initial distribution $s(x)$ and Markov dynamics $q(x_2, t_2|x_1, t_1), a \le t_1 < t_2 \le b$, and given begin and end distribution $p_a(x)$ and $p_b(z)$, compute the marginal distribution of most likely trajectories from a to b for any intermediate time $a \le t \le b$.

Link between control theory and large deviations

Consider first the simpler Brownian bridge problem: $p_a(x') = \delta(x' - x), p_b(z') = \delta(z' - z)$.

The probability of y at an intermediate time $a < t < b$ is given by Bayes rule:

$$q(y,t|x,a;z,b) = \frac{q(z,b|y,t)q(y,t|x,a)}{q(z,b|x,a)}$$

$$q(z,b|x,a) = \int dyq(z,b|y,t)q(y,t|x,a)$$

If x, z from $p(x,z)$ with marginals $p_a(x)$ and $p_b(z)$, then

$$q(y,t|p_a;p_b) = \int dxdzq(y,t|x,a;z,b)p(x,z)$$

How to compute $p(x,z)$?

Link between control theory and large deviations

Imagining a cloud if N particles with distribution $s(x)$ at time a and move according to a Markov process $q(z, b|x, a)$.

We discretize space in partitions Δx_k and Δz_l and denote $\Gamma_{kl} = \#$ particles leaving cell Δx_k and arriving in cell Δz_l.

The probability for an individual particle trajectory is $s_k q_{kl} \Delta x_k \Delta z_l$ and Γ follows a multinomial distribution

$$P(\Gamma) = N! \prod_{kl} \frac{(s_k q_{kl} \Delta x_k \Delta z_l)^{\Gamma_{kl}}}{\Gamma_{kl}!} \qquad s_k = s(x_k), \quad q_{kl} = q(z_l, b|x_k, a)$$

Define $p_{kl} \Delta x_k \Delta z_l = \Gamma_{kl}/N$, then

$$\lim_{N \to \infty} P(\Gamma) = \exp\left(-N KL(p||sq)\right)$$

Link between control theory and large deviations

$KL(p||sq)$ must be minimized wrt p subject to the constraints

$$\sum_x p(x,z) = p_b(z) \qquad \sum_z p(x,z) = p_a(x)$$

The solution is $p(x,z) = q(z,b|x,a)\hat{\phi}(x,a)\phi(z,b)$.

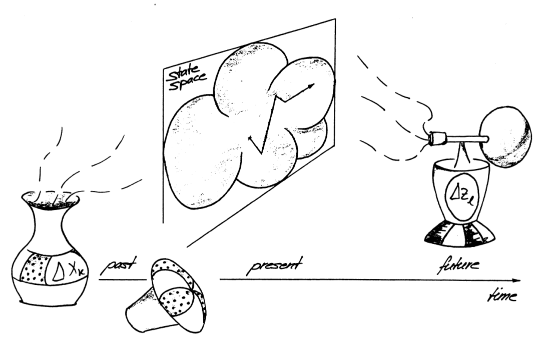
The solution for the Schrödinger bridge thus becomes

$$\begin{aligned} q(y,t|p_a;p_b) &= \hat{\phi}(y,t)\phi(y,t) \\ \hat{\phi}(y,t) &= \int dx q(y,t|x,a)\hat{\phi}(x,a) \\ \phi(y,t) &= \int dz q(z,b|y,t)\phi(z,b) \end{aligned}$$

The (mixed) boundary conditions for $\phi, \hat{\phi}$ are

$$p_a(x) = \hat{\phi}(x,a)\phi(x,a) \qquad p_b(x) = \hat{\phi}(x,b)\phi(x,b)$$

Link between control theory and large deviations



The solution for the Schrödinger bridge thus becomes

$$\begin{aligned} q(y,t|p_a;p_b) &= \hat{\phi}(y,t)\phi(y,t) \\ \hat{\phi}(y,t) &= \int dx q(y,t|x,a)\hat{\phi}(x,a) \\ \phi(y,t) &= \int dz q(z,b|y,t)\phi(z,b) \end{aligned}$$

The (mixed) boundary conditions for $\phi, \hat{\phi}$ are

$$p_a(x) = \hat{\phi}(x,a)\phi(x,a) \qquad p_b(x) = \hat{\phi}(x,b)\phi(x,b)$$

KL control theory

x denotes state of the agent and $x_{1:T}$ is a path through state space from time $t = 1$ to T.

$q(x_{1:T}|x_0)$ denotes a probability distribution over possible future trajectories given that the agent at time $t = 0$ is is state x_0, with

$$q(x_{1:T}|x_0) = \prod_{t=0}^{T} q(x_{t+1}|x_t)$$

$q(x_{t+1}|x_t)$ implements the allowed moves.

$R(x_{1:T}) = \sum_{t=1}^{T} R(x_t)$ is the total cost when following path $x_{1:T}$.

The KL control problem is to find the probability distribution $p(x_{1:T}|x_0)$ that minimizes

$$C(p|x_0) \quad = \quad \sum_{x_{1:T}} p(x_{1:T}|x_0) \left(\log \frac{p(x_{1:T}|x_0)}{q(x_{1:T}|x_0)} + R(x_{1:T}) \right) = KL(p||q) + \langle R \rangle_p$$

KL control theory

$p(x_{1:T}|x_0)$ and $q(x_{1:T}|x_0)$ distributions over trajectories.

Given q, find p that minimizes

$$C(p|x_0) = KL(p||q) - \langle R \rangle_p$$

The solution and the optimal control cost are

$$\begin{aligned} p(x_{1:T}|x_0) &= \frac{1}{Z(x_0)} q(x_{1:T}|x_0) \exp\left(R(x_{1:T})\right) \\ C &= -\log Z(x_0) \\ Z(x_0) &= \sum_{x_{1:T}} q(x_{1:T}|x_0) \exp\left(R(x_{1:T})\right) \end{aligned}$$

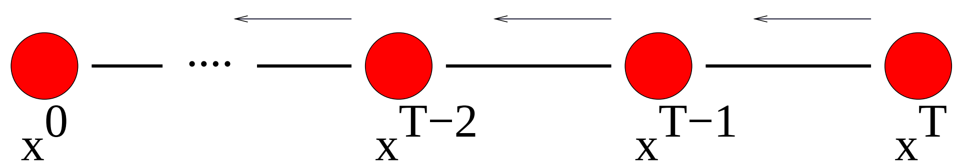
NB: $Z(x_0)$ is an integral over paths.

KL control theory

The optimal control at time $t = 0$ is given by

$$p(x_1|x_0) = \sum_{x_{2:T}} p(x_{1:T}|x_0) \propto q(x_1|x_0)\exp(R(x_1))\beta_1(x_1)$$

with $\beta_t(x)$ the backward messages.



$$\begin{aligned}\beta_T(x_T) &= 1 \\ \beta_{t-1}(x_{t-1}) &= \sum_{x_t} q(x_t|x_{t-1})\exp(R(x_t))\beta_t(x_t)\end{aligned}$$

KL control theory

The control computation is 'reduced' to a (graphical model) inference problem.

Dynamics: $p^t_{xy}(\pi)$ Cost: $C(\pi^{0:T}) = \langle R \rangle$	$\rightarrow$ DP $\rightarrow$	Bellman Equation
$\downarrow$		$\downarrow$
restricted class		approximate J
$\downarrow$		$\downarrow$
Free dynamics: q^t_{xy} $C = KL(p\|\|q \exp(R))$	$\rightarrow$ approx inference $\rightarrow$	Optimal π

Optimal solution is a Gibbs distribution

$$p(x^{1:T}|x^0) = \frac{1}{Z} q(x^{1:T}|x^0) \exp(R(x^{0:T}))$$

Link to continuous path integral formulation

The previous continuous path integral control can be obtained as a special case of the KL control formulation.

$$
\begin{aligned}
p(x_{t+dt}|x_t, u_t) &= \mathcal{N}(x_{t+dt}|x_t + f(x_t, t)dt + u_t dt, \nu) \\
q(x_{t+dt}|x_t) &= \mathcal{N}(x_{t+dt}|x_t + f(x, t)dt, \nu) \\
C(p|x_0) &= KL(p|q) - \langle R \rangle = \sum_{x^{dt:T}} p(x^{dt:T}|x^0) \left(\sum_{t=dt}^{T} \frac{1}{2} u_t^T \nu^{-1} u_t - R(x_t) \right)
\end{aligned}
$$

Average cost KL control (Todorov 2006)

When $T \to \infty$ and q ergodic the backward message recursion

$$\beta_{t-1}(x_{t-1}) \quad = \quad \sum_{x_t} H(x_{t-1}, x_t)\beta_t(x_t) \qquad H(x,y) = q(y|x)\exp(R(y))$$

becomes the computation of the Perron-Frobenius eigen pair $(\beta(\cdot), \lambda)$:

$$H\beta = \lambda\beta \qquad H(x,y) = q(y|x)\exp(R(x))$$

The optimal control satisfies

$$\begin{aligned} p(y|x) &= q(y|x)\exp(R(x))\frac{\beta(y)}{\lambda\beta(x)} \\ C &= -\log\lambda \\ J(x) &= -\log\beta(x) \end{aligned}$$

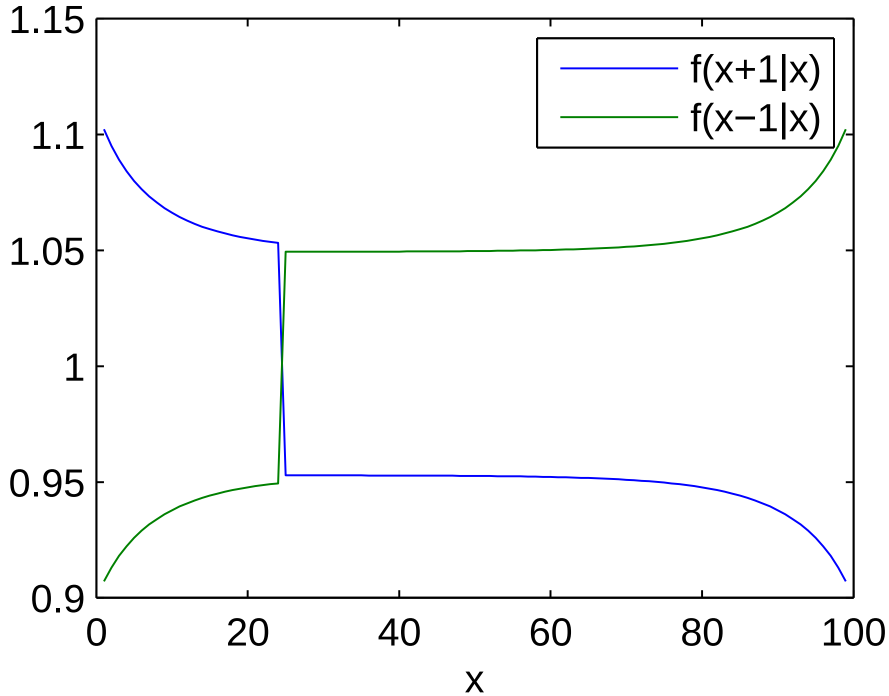
Detailed balance

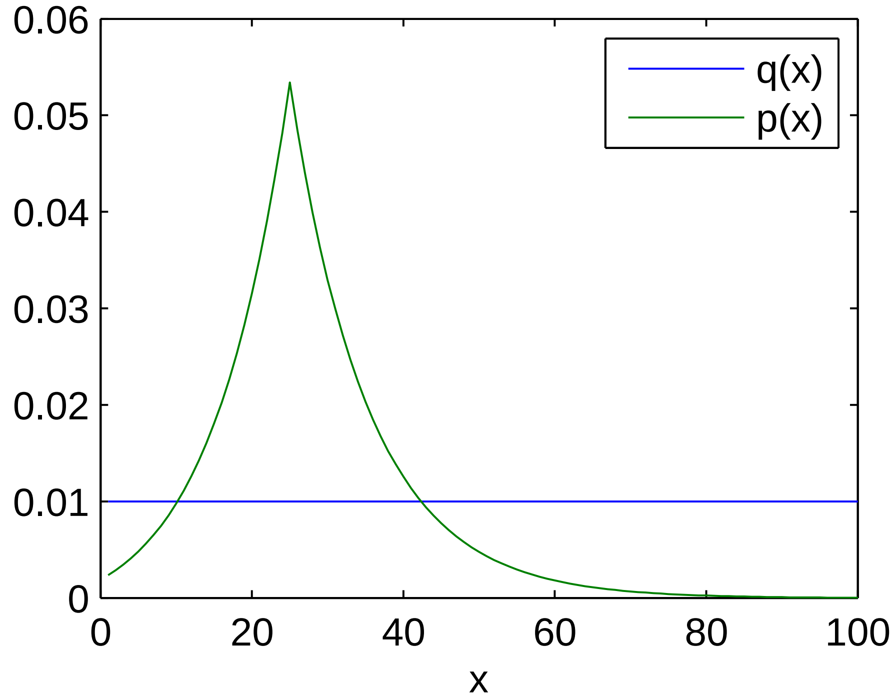
When the uncontrolled dynamics satisfies DB with equilibrium distribution $q(x)$, the optimal control solution also satisfies DB with

$$p(x) \propto q(x) v_1^2(x)$$

Particle in a box

Consider a one dimensional grid $x = 1, 2, \ldots, 100$. Uncontrolled dynamics $q(x \pm 1|x)$. Cost $R = 1$ at the walls and $R = -1$ at target. $p(y|x) = \delta_{y,x} + f(y|x)dt$





KL-learning [Bierkens, Kappen 2012]

- Goal: find Perron-Frobenius solution $Hz = \lambda z$, with $H = [q(y|x)\exp(-R(x))]$, while stepping through state space according to q and observing incurred cost.

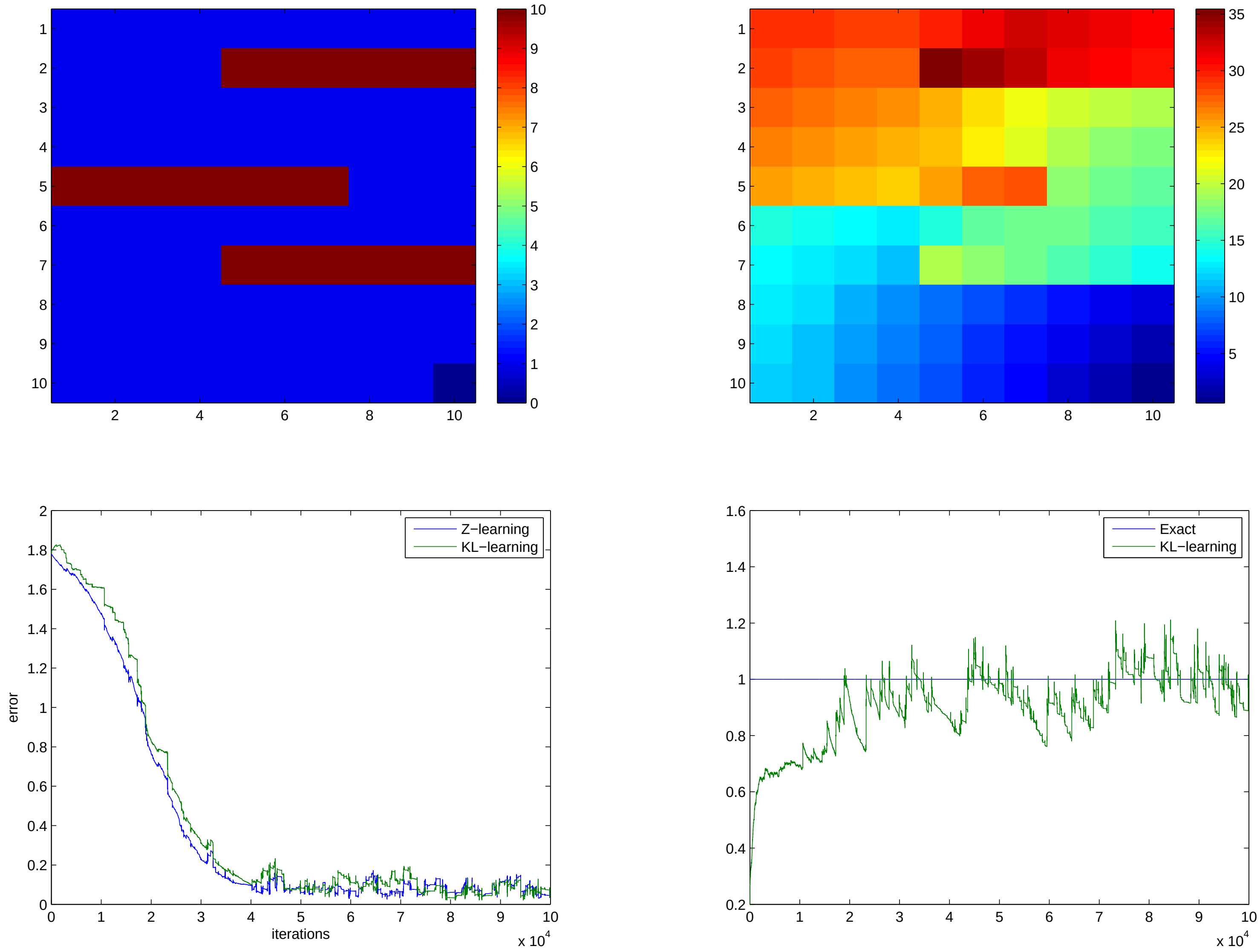
- Algorithm (KL-learning):

  $z \leftarrow (1/n, \ldots, 1/n)$, $\lambda > 0$, $x \leftarrow$ any state
  for $m = 1 : M$ **do**
  $\quad y \leftarrow$ independent draw from $q(\cdot|x)$
  $\quad \Delta \leftarrow \exp(-R(x))z(y)/\lambda - z(x)$
  $\quad z(x) \leftarrow z(x) + \gamma\Delta$
  $\quad \lambda \leftarrow \lambda + \gamma\Delta$
  $\quad x \leftarrow y$
  end for

- Invariants: $z > 0$, $\lambda = ||z||_1$.

Generalization of z-learning (Todorov) to $\lambda \neq 1$

Numerical experiment

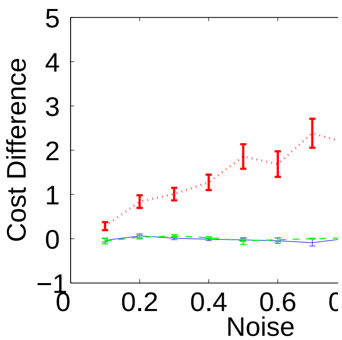


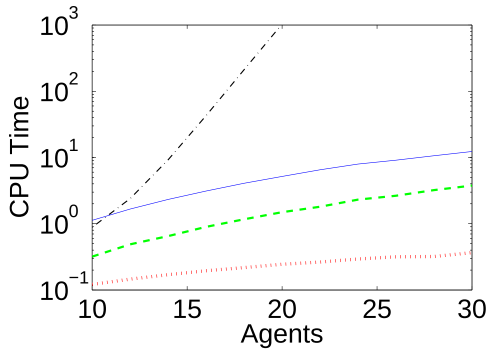
Summary and discussion

- QM does not (seem to) correspond to a 'useful' class of control problems

Summary and discussion

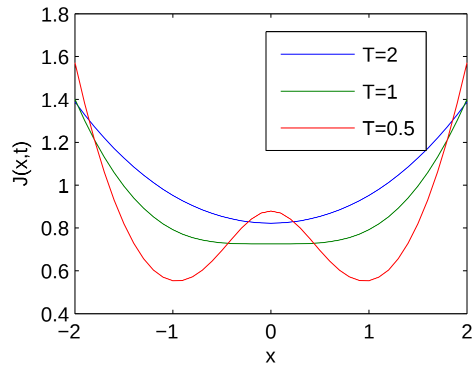
- QM does not (seem to) correspond to a 'useful' class of control problems
- Control as inference links control to machine learning and statistical physics
  - efficient computational methods





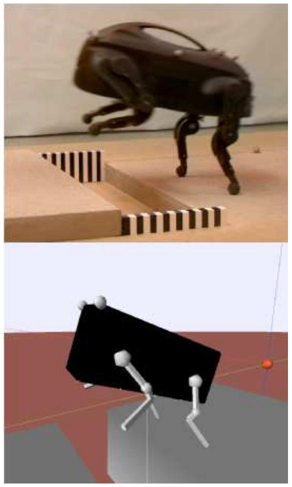
Summary and discussion

- QM does not (seem to) correspond to a 'useful' class of control problems
- Control as inference links control to machine learning and statistical physics
  - efficient computational methods
  - insight in the role of noise: phase transitions (delayed choice)



Summary and discussion

- QM does not (seem to) correspond to a 'useful' class of control problems
- Control as inference links control to machine learning and statistical physics
  - efficient computational methods
  - insight in the role of noise: phase transitions (delayed choice)
  - favorable comparison with state-of-the-art RL methods in robotics (Theodorou 2010-2012)

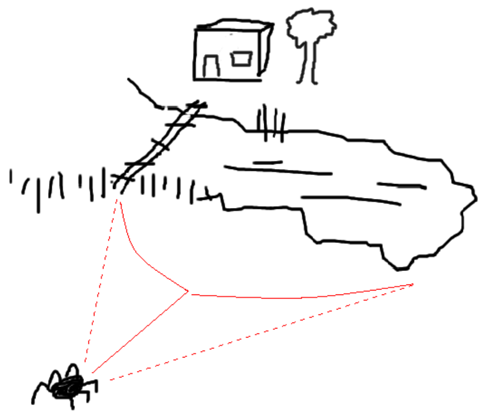


Summary and discussion

- QM does not (seem to) correspond to a 'useful' class of control problems
- Control as inference links control to machine learning and statistical physics
  - efficient computational methods
  - insight in the role of noise: phase transitions (delayed choice)
  - favorable comparison with state-of-the-art RL methods in robotics (Theodorou 2010-2012)
- Large deviations motivate KL control and possible generalizations

Further reading

http://www.snn.ru.nl/~bertk/
http://www.snn.ru.nl



Some open questions

Partially observable control problems correspond to non-linear controlled diffusions.

Does there exist a class of partially observed control problems that can be linearized?

Some open questions

In path integral control theory the optimal cost to go is a free energy:

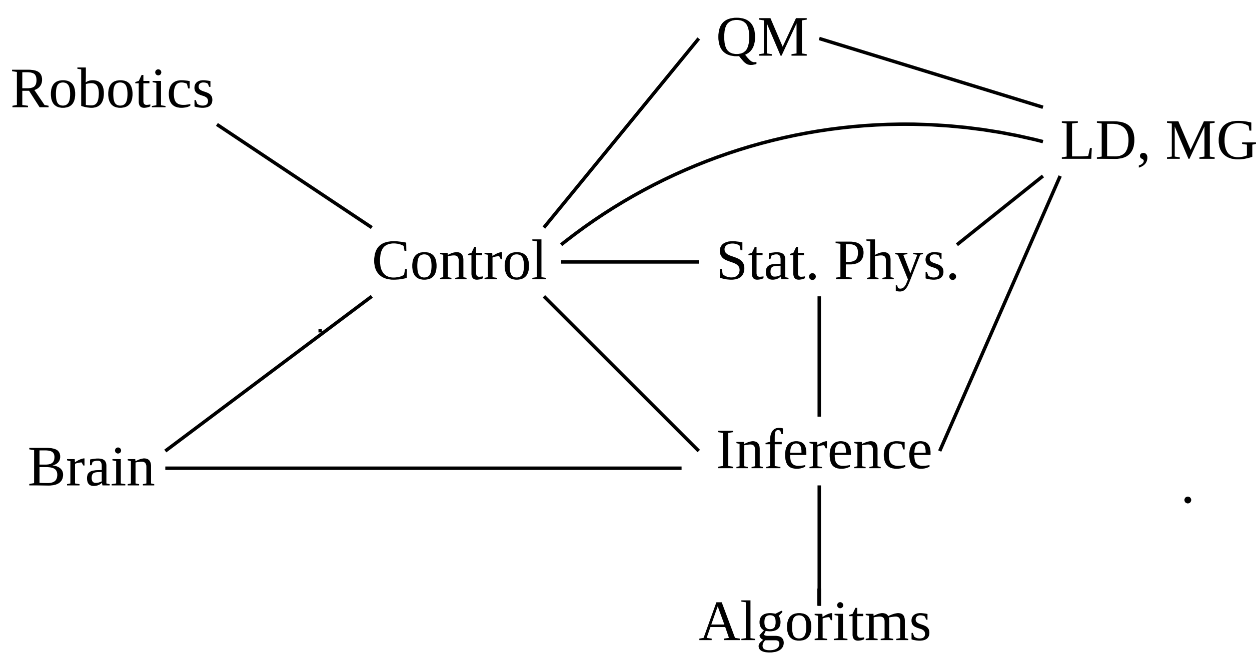
$$J(x,t) = \int dQ(\tau|x,t) \exp\left(-S(\tau|x,t)/\lambda\right)$$

The Jarzynski equation relates the free energy difference of two states to a path integral over the work

$$\left\langle \exp\left(-\int dt dW(t)\right)\right\rangle = \exp(-\Delta F)$$

Are these two statements related?

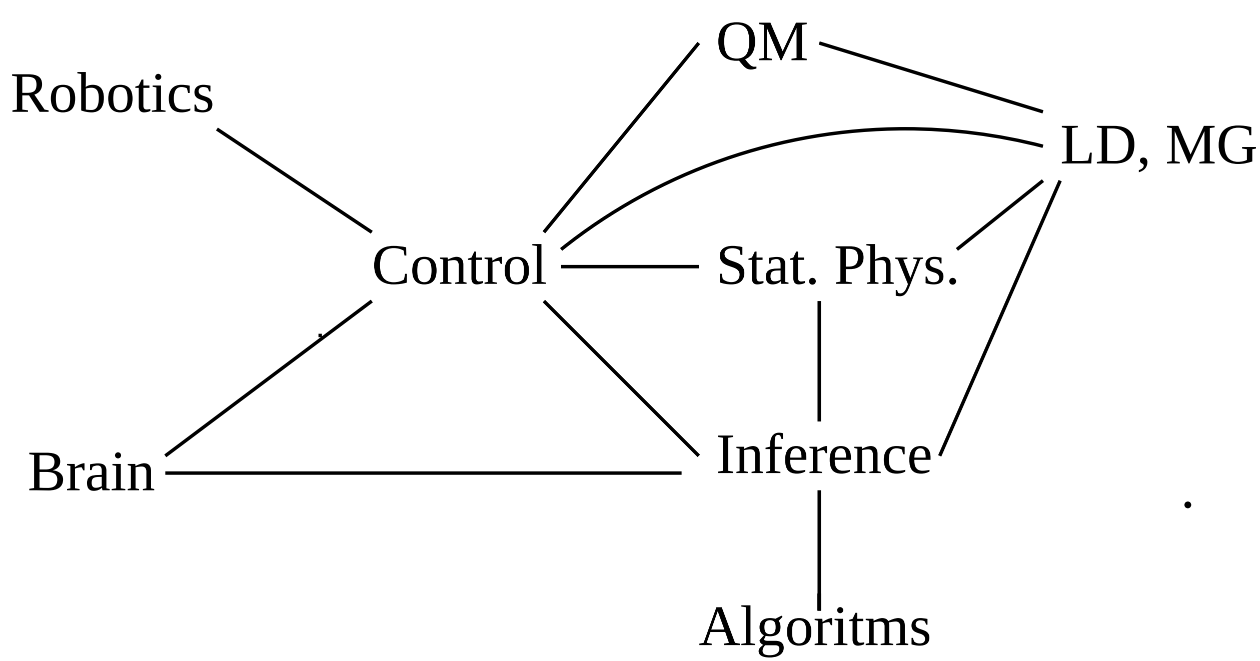
The statistical physics of control and inference



Stat. Phys, Inference, Algorithms:

Mezard, Zecchina, Shah, Parrilo, Dall'Asta, Nakano

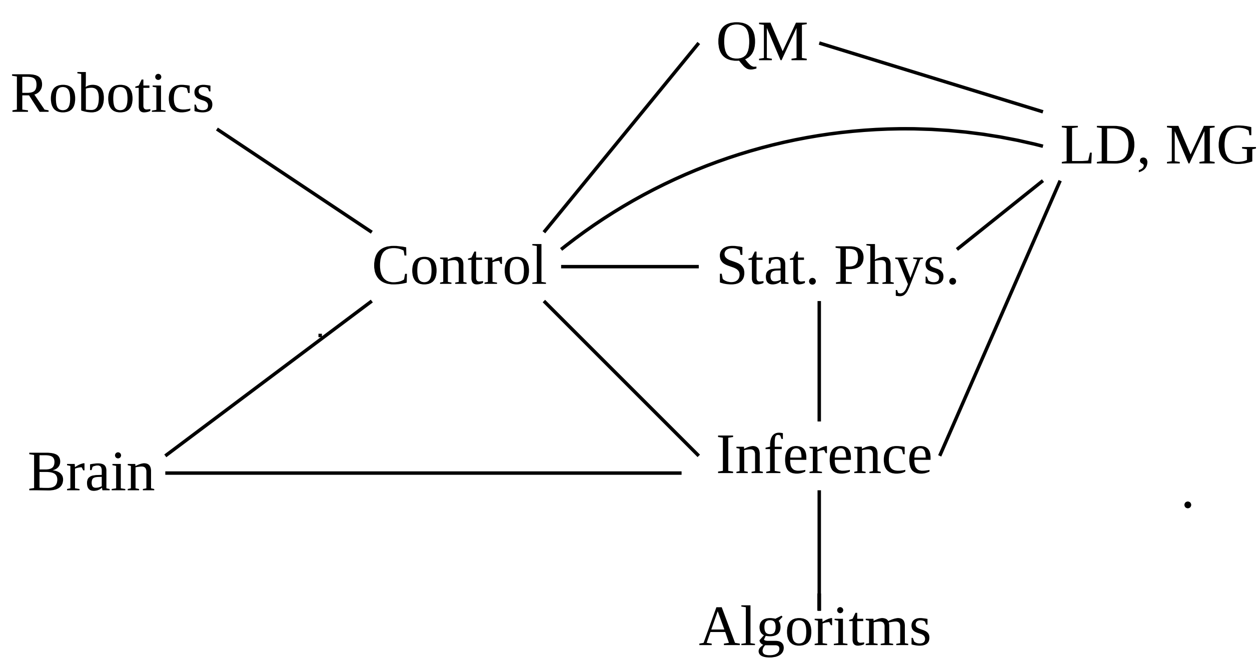
The statistical physics of control and inference



Quantum Mechanics and control:

Guerra, Lloyd, Maassen

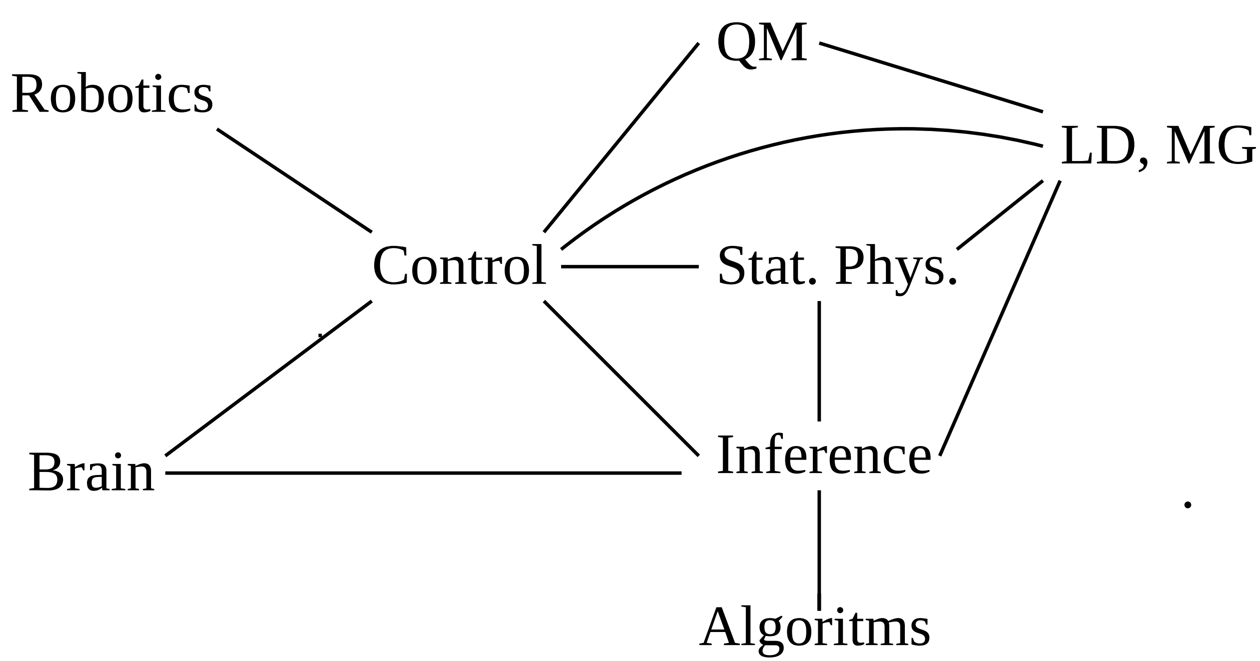
The statistical physics of control and inference



Control, statistical physics and LD:

Landim, Sandberg, Delvenne, van Handel, Chernyak, Chertkov, Parrondo, Seldin, Wiegerinck

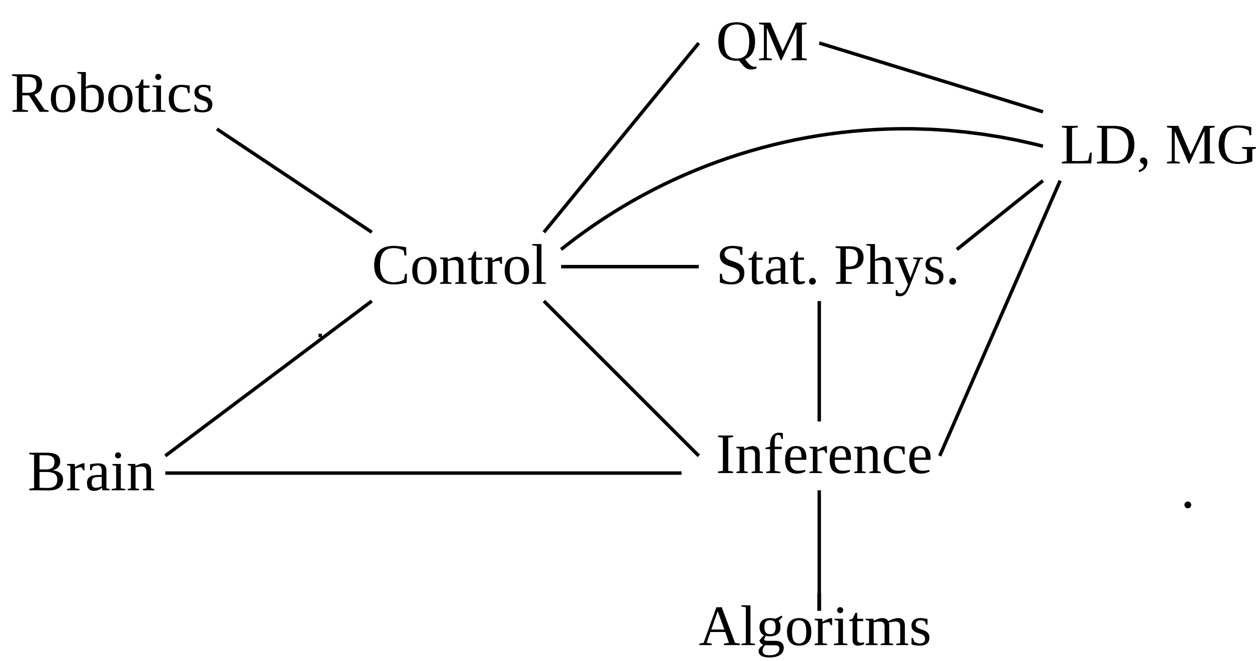
The statistical physics of control and inference



Relation between control and inference:

Mitter, Bierkens, Hurtado, Brockett

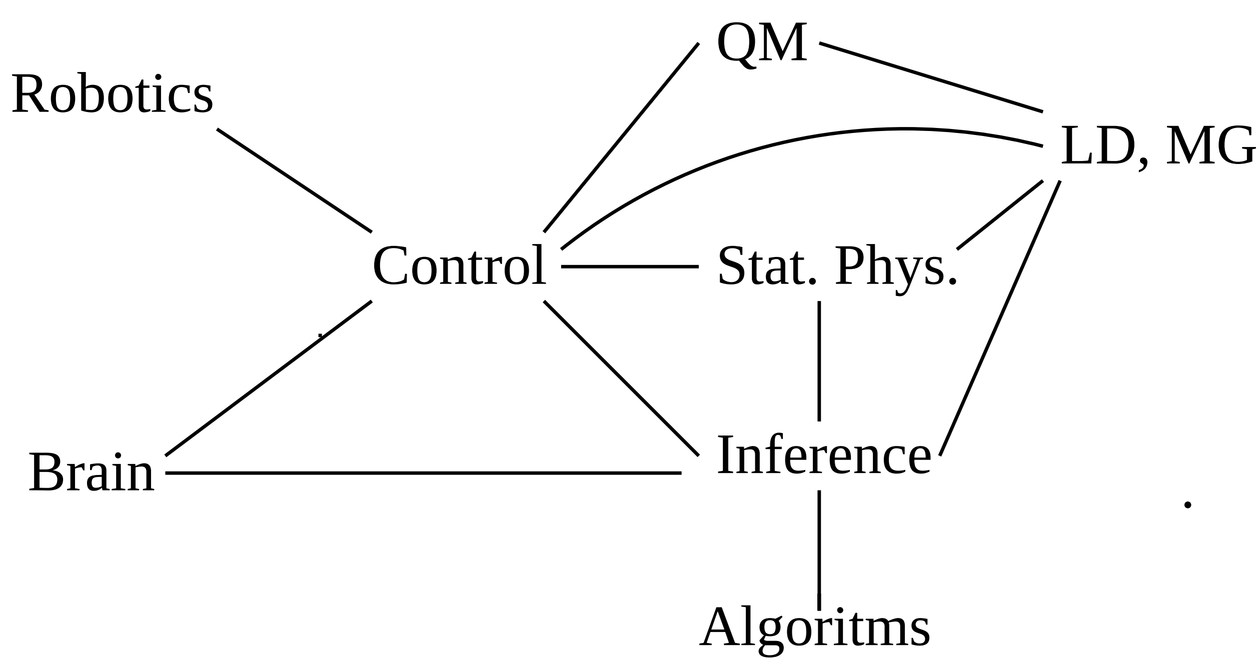
The statistical physics of control and inference



Relation to biology:

Friston, Tishby, Todorov, Braun

The statistical physics of control and inference



Application in robotics and control:

Theodorou, Morimoto, Satoh, Dvijotham