Numerical Procedures for Dynamic Programming

A. N. Daryin and A. B. Kurzhanski

Moscow State (Lomonosov) University
Faculty of Computational Mathematics and Cybernetics

Dynamical Systems: Stability, Control, Optimization
September 30, 2008, Minsk, Belarus

Outline

Introduction

- Dynamic programming (DP) is a key tool in understanding and solving feedback control problems.
- Direct application of DP methods may lead to high computation load.
- But: avoiding DP due to computational complexity is an "infantile disorder".
- Instead: modify solutions to require less computation. Here — ellipsoidal approximations.

The Ellipsoidal Approximations

Definition

An ellipsoid

$$\mathscr{E}(q, Q) = \{x \mid \|x - q\|_Q^2 \leq 1\}$$

q — center Q — configuration matrix $normx_Q^2 = \langle x, Qx \rangle$

Ellipsoidal approximation = approximating value function by quadratic form:

$$V(t, x) \leq (\text{or } \geq) \langle x - q, Q(x - q) \rangle + k(t).$$

The Ellipsoidal Toolbox

Ellipsoidal toolbox (A. A. Kurzhanskiy and P. Varaiya):

http://code.google.com/p/ellipsoids/

ET is a part of **Multi-Parametric Toolbox** (MPT)

http://control.ee.ethz.ch/~mpt/

The Comparison Principle

Deriving ellipsoidal approximations:

Inductive approach
based on ellipsoidal calculus.

Deductive approach
directly from HJB equation, basing on Comparison Principle.

The Comparison Principle

For a system of type

$$\dot{x} = f(t, x, u), \quad u(t) \in \mathscr{P}(t),$$

the solution $V_0(t, x)$ to the HJB equation

$$V_{0t} + H_0(t, x, V_{0x}) = 0,$$

produces backward reach set $W[t] = W(t, \vartheta, \mathscr{M})$ as its level set.

Here

$$H_0(t, x, p) = \min\{(p, f(t, x, u)) \mid u \in \mathscr{P}(t)\}.$$

The Comparison Principle

Theorem

Suppose that given are $H_+(t,x,p)$ *and* $w(t,x) \in C_1$, $\mu(t) \in L$, *which satisfy the inequalities*

$$H(t,x,p) \leq H_+(t,x,p), \quad \forall \{t,x,p\},$$

$$w_t + H_+(t,x,w_x) \leq \mu(t).$$

Then there exists an upper estimate

$$X_+[t] \supseteq X[t],$$

where

$$X_+[t] =$$

$$= \left\{ x \;\middle|\; w(t,x) \leq \int_{t_0}^{t} \mu(s)ds + \max_{x \in X^0} w(t_0,x) \right\}.$$

Backward Reach Set Approximation

A similar theorem is true for the backward reach sets.

Theorem

Suppose there exists function $H_-(t,x,p)$ and functions $w^0(t,x) \in C_1$, $\nu(t) \in L_1$ which satisfy inequalities

$$H_0(t,x,p) \geq H_-(t,x,p), \quad \forall \{t,x,p\},$$

$$w_t^0 + H_-(t,x,w_x^0) \geq \nu(t).$$

Then there exists an upper estimate

$$W_+[t] \supseteq W[t],$$

where

$$W_+[t] =$$

$$= \left\{ x \,\middle|\, w^0(t,x) \leq \max_{x \in \mathscr{M}} w^0(t_1,x) - \int_{t_0}^{t} \nu(s) ds \right\}.$$

Internal Approximations

For **internal** approximations of sets $W[t]$ we have to approximate **from above** the value function $V_0(t, x)$ which solves the same HJB equation with boundary condition $V(t_0, x)$.

Then the following assertion is true.

Internal Approximations

Theorem

Suppose that there exists a function $h(t,x,p)$,

$$h(t,x,p) \leq H(t,x,p), \quad \forall t,x,p,$$

together with a continuously differentiable function $\psi(t,x)$ *which satisfies equation*

$$\psi_t + h(t,x,\psi_x) = 0, \quad \forall t \in [t_0, \vartheta]$$

with boundary condition $\psi(t_0,x) = V(t_0,x)$.
Then the next inclusion is true

$$W_-[t] = \{x \mid \psi(t,x) \leq 0\} \subseteq W[t],$$

where $W[t] = \{x \mid V_0(t,x) \leq 0\}$.

The Impulse Control Problem

$$\dot{x}(t) = A(t)x(t) + B(t)u(t)$$

$t \in [t_0, t_1]$ — fixed time interval

Problem (1, a Mayer–Bolza Analogy)

Minimize $J(U(\cdot)) = \operatorname*{Var}_{[t_0,t_1]} U(\cdot) + \varphi(x(t_1 + 0))$

over $U(\cdot) \in BV[t_0, t_1]$ *with* $x(t)$ *generated by control input*

$$u(t) = \frac{dU}{dt}$$

starting from $x(t_0 - 0) = x_0$.

The Impulse Control Problem

Known result (N. N. Krasovski [1957], L. W. Neustadt [1964]):

$$u(t) = \sum_{i=1}^{n} h_i \delta(t - \tau_i)$$

Important particular case: $\varphi(x) = \mathscr{I}(x \mid \{x_1\})$
— steer from x_0 to x_1 on $[t_0, t_1]$.

$$\mathscr{I}(x \mid A) = \begin{cases} 0, & x \in A; \\ +\infty, & x \notin A. \end{cases}$$

The Value Function

Definition

The minimum of $J(U(\cdot))$ with *fixed* initial position $x(t_0 - 0) = x_0$ is called the value function:

$$V(t_0, x_0) = V(t_0, x_0; t_1, \varphi(\cdot)).$$

How to find the value function?

- Integrate the HJB equation.
- An explicit representation (convex analysis).
- Ellipsoidal approximation (comparison principle).

The Dynamic Programming Equation

The value function $V(t, x; t_1, \varphi(\cdot))$ satisfies the Principle of Optimality

$$V(t_0, x_0; t_1, \varphi(\cdot)) = V(t_0, x_0; \tau, V(\tau, \cdot; t_1, \varphi(\cdot))), \quad \tau \in [t_0, t_1]$$

The value function it is the solution to the Hamilton–Jacobi–Bellman variational inequality:

$$\min \{H_1(t, x, V_t, V_x),\ H_2(t, x, V_t, V_x)\} = 0,$$

$$V(t_1, x) = V(t_1, x; t_1, \varphi(\cdot)).$$

$$H_1 = V_t + \langle V_x, A(t)x \rangle, \quad H_2 = \min_{u \in S_1} \langle V_x, B(t)u \rangle + 1 = -\left\| B^T(t) V_x \right\| + 1.$$

The Control Structure

$$H_1(t,x)=0 \longleftarrow (t,x) \longrightarrow H_2(t,x)=0$$

From $H_1(t,x)=0$ ↓

wait

$$dU(t)=0$$

From $H_2(t,x)=0$ ↓

choose jump direction

$$d=-B^T V_x$$

↓

choose jump amplitude

$$\min \alpha \geq 0 : H_1(t, x+\alpha d)=0$$

↓

jump

$$U(\tau)=\alpha \cdot d \cdot \chi(\tau - t)$$

The Explicit Formula

$$V(t_0, x_0) = \inf_{x_1 \in \mathbb{R}^n} \{\varphi(x_1) + \sup_{p \in \mathbb{R}^n} \frac{\langle p, x_1 - X(t_1, t_0)x_0 \rangle \parallel \|p\|_{[t_0,t_1]}}{\}}.$$

The value function is convex and its conjugate equals

$$V^*(t_0, p) = \varphi^*(X^T(t_0, t_1)p) + \mathscr{I}\left(X^T(t_0, t_1)p \mid \mathbb{B}_{\|\cdot\|_{[t_0,t_1]}}\right)$$

where $\|p\|_{[t_0,t_1]} = \left\|B^T(\cdot)X^T(t_1, \cdot)p\right\|_{C[t_0,t_1]}$ and $\partial X(t, \tau) = A(t)X(t, \tau)$, $X(\tau, \tau) = I$.

See (Daryin, Kurzhanski, and Seleznev, 2005).

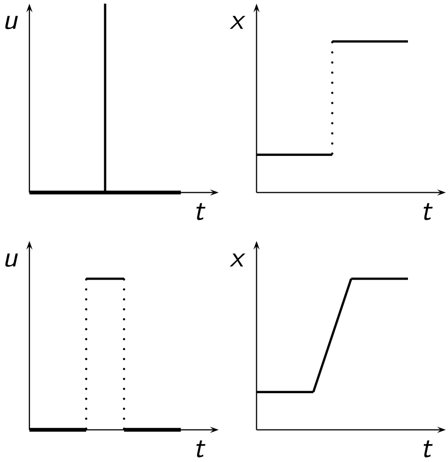
The Approximate DP Problem

We now approximate the nonstandard DP problem for impulse controls by a relatively standard problem with double constraints.

What for?

The fact is that "ideal" impulse controls, as taken in the mathematical sense, are not physically realizable whereas the approximations of impulses through "ordinary-type functions" are realizable.

The Approximate DP Problem



ideal scheme

real scheme

The Approximate DP Problem

Minimize integral

$$J_\mu(u(\cdot)) = \int_{t_0}^{t_1} \|u(t)\| dt + \phi(x(t_1)) \to \inf$$

over all $u(\cdot) \in L_1([t_0, t_1]; \mathbb{R}^m)$ due to equation

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \quad x(t_0) = x_0$$

under blue additional constraint

$$\|u(t)\| \leq \mu, \quad t \in [t_0, t_1]$$

The Value Function for the Approximate Problem

$$V_\mu(t_0, x_0) = \min\{J_\mu(u(\cdot))|x(t_0) = x_0\}$$

$$V_\mu(t_0, x_0) = V_\mu(t_0, x_0; t_1, \phi(\cdot))$$

Function V_μ is a classical solution of the next HJB equation (in the general case it is a generalized viscosity solution)

$$\frac{\partial V_\mu}{\partial t} + \min_{\|u\| \leq \mu} \left\{ \left\langle \frac{\partial V_\mu}{x}, A(t)x(t) + B(t)u \right\rangle - \|u\| \right\} = 0$$

$$V_\mu(t_1, x) = \phi(x)$$

Internal Ellipsoidal Approximation

Ellipsoidal approximation is derived through comparison principle:

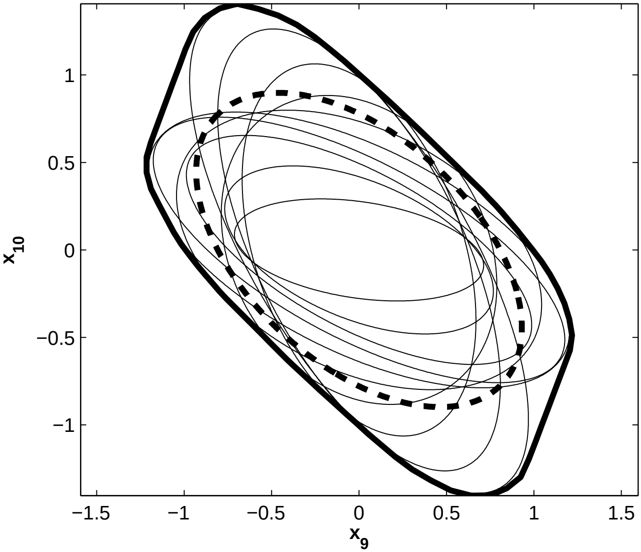
$$\mathscr{X}_\nu^-[t] = \mathscr{E}(0, (\nu - k(t))Z(t))$$

$$\begin{cases} \dot{Z} & = AZ + ZA^T - \eta(t)BB^T \\ \dot{k} & = -\frac{1}{4}\eta(t) \end{cases} \qquad \begin{cases} Z(t_1) & = 0 \\ k(t_1) & = 0 \end{cases}$$

Here $\eta(t) \geq 0$ is a parameter function

$$\mathscr{X}_\nu[t] = \mathrm{cl} \bigcup_{\nu(\cdot)} \mathscr{X}_\nu^-[t]$$

Internal Ellipsoidal Approximation



The Double Constraints Problem

Problem

Find backward reach set (solvability domain)

$$\mathscr{W}[t_0] = \mathscr{W}(t_0, t_1, \mathscr{M}, k_0)$$

for linear system

$$\dot{x}(t) = A(t)x(t) + B(t)u(t)$$

under soft bound

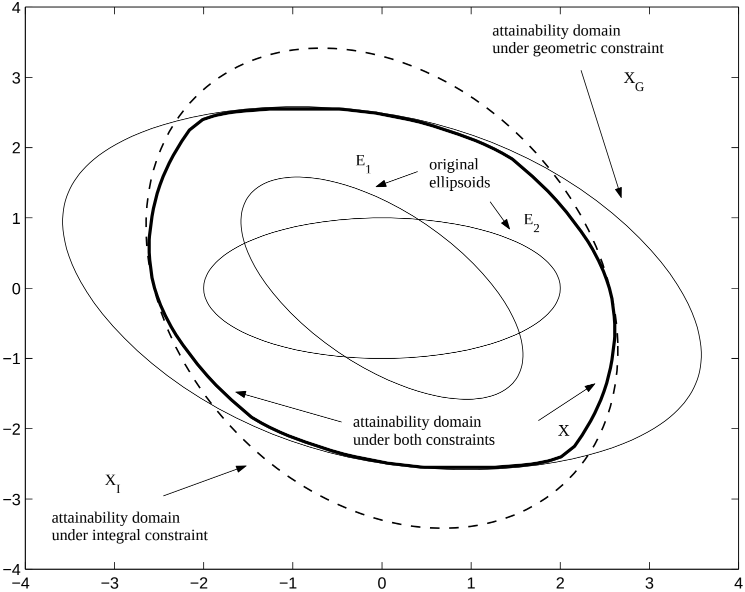
$$\int_{t_0}^{t_1} \|u(t)\|_N^2 dt \leq k(t_0) = k_0 > 0$$

and hard bound

$$u(t) \in \mu\mathscr{E}(0, Q)$$

given target set $\mathscr{M} = \mathscr{E}(m, M)$.

Double Constraints — An Illustration



The Value Function

The Value Function

The value function is defined as

$$V(t,x,k) = \min_{u(\cdot)}\{d^2(x[t_1],\mathscr{M}) + ((k[t_1])_-)^2 \mid x[t] = x, k[t] = k\}.$$

Here

$$k(t) = k_0 - \int_{t_0}^{t} \|u\|_N^2 dt,$$

$$\dot{k}(t) = -\|u\|_N^2.$$

$$(k)_- = \min\{0, k\}.$$

The Dynamic Programming Equation

The Hamilton–Jacobi–Bellman equation

$$V_t + \min_u \Big\{ \langle V_x, A(t)x + B(t)u \rangle - V_k \chi[t] \|u\|_N^2 \;\Big|\; (1-\chi[t])u \in \mu \mathscr{E}(0, Q) \Big\} = 0$$

under boundary condition $V(t_1, x, k) = d(x^2, \mathscr{M}) + ((k)_-)^2$.

Here

$$\chi[t] = \chi(t, x, k) = \begin{cases} 0, & \text{hard bound active,} \\ 1, & \text{hard bound inactive.} \end{cases}$$

The Hamiltonian

The Hamiltonian

$$H(t,x,k,\xi,\varkappa) = (1-\chi[t])H_0(t,x,\xi) + \chi[t]H_1(t,x,k,\xi,\varkappa).$$

For $\chi[t] = 0$

$$H_0(t,x,\xi) = \langle \xi, A(t)x \rangle - \left\langle \xi, BQB^T\xi \right\rangle^{\frac{1}{2}},$$

For $\chi[t] = 1$

$$H_1(t,x,k,\xi,\varkappa) = \min_u \{\langle \xi, A(t)x + B(t)u \rangle - \varkappa \|u\|_N^2\}.$$

External Approximation

We approximate the value function from below by

$$w(t,x,k) = \left\langle x - x^*(t), \mathscr{K}_+^{-1}(t)(x - x^*(t))\right\rangle + \chi(t)k - 1.$$

External Approximation

$$\mathscr{W}[t] \subseteq \mathscr{E}(x^*, \mathscr{K}_+(t))$$

$$\dot{\mathscr{K}}_+ = A(t)\mathscr{K}_+ + \mathscr{K}_+ A^T(t) + \pi(t)\mathscr{K}_+ - \pi^{-1}(t)B(t)Q(T)B^T(t)$$

$$\dot{x}^*(t) = A(t)x^*(t)$$

$$x^*(t_1) = m, \quad \mathscr{K}_+(t) = M.$$

Internal Approximation

Here we approximate $V(t,x,k)$ by quadratic form from above.

Internal Approximation

$$\mathscr{W}[t] = \mathscr{E}(x^*(t), \mathscr{K}_-(t))$$

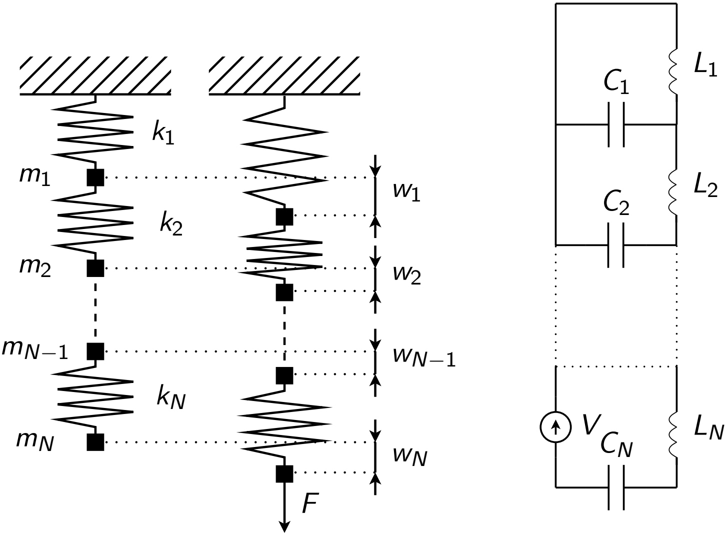
$$\dot{\mathscr{K}}_- = A(t)\mathscr{K}_- + \mathscr{K}_- A^T(t) + r^{-1}(t)(\mathscr{K}_- S(t)\mathbf{B}(T) + \mathbf{B}(t)S^T(t)\mathscr{K}_-)$$

$$\mathbf{B}(t) = (B(t)Q(T)B^T(t))^{\frac{1}{2}}.$$

$$\dot{x}^*(t) = A(t)x^*(t).$$

$r(t) > 0$ is a **tuning parameter**.
$S(t)$ is an orthogonal matrix function. (for case $\chi[t] = 0$.)

Examples — Oscillating Systems



The Oscillating System Equations

$$\begin{cases} m_1 \ddot{w}_1 = k_2(w_2 - w_1) - k_1 w_1 \\ m_i \ddot{w}_i = k_{i+1}(w_{i+1} - w_i) - k_i(w_i - w_{i-1}) \\ m_\nu \ddot{w}_\nu = k_{\nu+1}(w_{\nu+1} - w_\nu) - k_\nu(w_\nu - w_{\nu-1}) + u(t) \\ m_N \ddot{w}_N = -k_N(w_N - w_{N-1}) \end{cases}$$

- $w_i = w_i(t)$ — displacements from the equilibrium
- m_i — masses of the loads
- k_i — stiffness coefficients
- $u(t) = \frac{dU}{dt}$ — impulse control $(U \in BV)$
- Dimension is $2N$ (40 for 20 springs).

$N \to \infty$: the string equation

$$\rho(\xi) w_{tt}(t,\xi) = [Y(\xi) w_\xi(t,\xi)]_\xi \,, \quad t > t_0, \quad 0 < \xi < L$$

$$w(t,0) = 0, \quad w_\xi(t,L) = u(t)/Y(L), \quad t \geq t_0$$

$$w(t_0,\xi) = w^0(\xi), \quad w_t(t_0,\xi) = \dot{w}^0(\xi), \quad 0 \leq \xi \leq L$$

- $w(t,\xi)$ — displacement from the equilibrium
- $u(t) = \frac{dU}{dt}$ — impulse control
- $\rho(\xi)$ — mass density
- $Y(\xi)$ — Young modulus

The Oscillating System

Normalized matrix form:

$$dx(t) = Ax(t)dt + B\,dU(t)$$

$$x(t) = \begin{pmatrix} w(t) \\ \dot{w}(t) \end{pmatrix} \quad w(t) = \begin{pmatrix} w_1(t) \\ \vdots \\ w_N(t) \end{pmatrix}$$

This system is **completely controllable**.

References

- *Kurzhanski A. B.* On problems of control synthesis under incomplete information. Vestnik MGU, 2005.
- *Bellman R., Dreyfus S.* Applied Dynamic Programming. Princeton Univ. Press, 1962.
- *Daryin A. N., Kurzhanski A. B.* Dynamic Programming for Problems of Control Synthesis // Proceedings of "Control Problems and Applications (technics, industry, economics)". Minsk, 2005.
- *Krasovski N. N.* On a problem of optimal regulation // Prikl. Math. & Mech. 1957. V. 21. N. 5. P. 670–677.
- *Kurzhanski A. B.* Comparison principle for Hamilton–Jacobi equations of control theory // Procedings of IMM UrO RAN. 2006. V. 12. P. 173–183.
- *Kurzhanskiy A.A., Varaiya P.* Ellipsoidal toolbox. `http://www.eecs.berkeley.edu/~akurzhan/ellipsoids/`, 2005.