On pathwise approach to optimal switching problems

Juri Hinz[1]

[1]UTS

December 2015

Research with Alex Novikov

- ARC Grant

  Risk Assessment of Climate Change Mitigation Measures

- Proposed research

  New concepts and applications of high-dimensional stochastic control

New Concepts:

- algorithmic approach
- high-dimensional solution methods
- duality & pathwise diagnostics
- applications in artificial intelligence

Motivation: Tiger game

Motivation: Tiger game

- behind one door is a tiger, behind the other a present
- open wrong door (tiger behind) costs \$100
- open correct door (present behind) gives \$10
- one can listen for \$1, but listening may give wrong observation, say with probability $1/3$
- upon a door is opened, tiger and present switch randomly, game starts again
- game played at times $0, \ldots, T$.

Such problems (POMDPs) are important in artificial intelligence

Motivation: Tiger game

Tiger game: wrong door, $100 penalty

Tiger game: correct door $10 reward

Motivation: Optimal asset liquidation

A broker must liquidate an asset within a fixed time

When submitting orders

- the time
- the size
- the order type

must be chosen optimally

Problem

At any time $t = 0, \ldots, T$, one knows

- number $p \in \mathbb{N}$ of asset units remaining
- current bid and ask prices

to decide on

- the size of the sell order
- the type of the sell order (limit/market)

limit order is valid for one step only

Problem

all randomness comes from the bid-ask spread, since price direction not predictable

revenue difference in order types is due to the current bid-ask spread

- **market order** sells with high probability at the current bid price
- **limit order** sells uncertain asset number at some higher (than current bid) price

Modeling as

- Discrete time stochastic control problems of specific type
- Efficient algorithms utilize linear state dynamics
- Solution diagnostics (duality of C. Rogers) is available

Target

- Solution (efficient implementation)
- Diagnostics (distance-to-optimality)

Stochastic switching with linear state dynamics

is about control problems whose state is $x = (p, z) \in P \times \mathbb{R}^d$

Discrete part is controlled Markov chain:

- Positions P (finite set)
- Actions A (finite set)
- Random jump $(p, a) \to \alpha(p, a) \in P$ with probability

$$\alpha^a_{p,p'} \in [0, 1], \qquad p, p' \in P, \quad a \in A$$

Continuous part is uncontrolled: $(Z_t)_{t=0}^T$ follows in $\mathbb{R}^d$

$$Z_{t+1} = W_{t+1} Z_t,$$

with independent disturbance matrices $(W_{t+1})_{t=0}^{T-1}$.

For asset liquidation, this would be

Discrete component:

finite set P of asset levels, actions A determine order type and size, whereas $\alpha^{a}_{p,p'}$ describes the level transition through the order a

Continuous component:

Spread size $(Z_t)_{t=0}^{T}$ follows Markov process.

This situation is frequent (Bermidian Put, Swing options, Storage valuation).

Efficient solutions and diagnostices

Optimal Stochastic Switching under Convexity Assumptions
SIAM Journal on Control and Optimization, 52(1), 2014

Using convex switching techniques for partially observable decision processes, Forthcoming in IEEE TAC

Algorithms for optimal control of stochastic switching systems
Forthcoming in TPA

For switching problems

stochastic control is as usual:

- Policy $\pi = (\pi_t)_{t=0}^{T-1}$ is a sequence of decision rules

$$\pi_t : P \times \mathbb{R}^d \to A \qquad (p, z) \mapsto \pi_t(p, z)$$

- Following π, one obtains for $t = 0, \ldots, T-1$

$$a_t^\pi := \pi_t(p_t^\pi, Z_t), \quad p_{t+1}^\pi := \alpha_{t+1}(p_t^\pi, a_t^\pi), \quad Z_{t+1} = W_{t+1} Z_t$$

started at $p_0^\pi = p_0$, $Z_0 = z_0 \in \mathbb{R}^d$.

Policy value

$$v_0^\pi(p_0, z_0) = \mathbb{E}\left(\sum_{t=0}^{T-1} r_t(p_t^\pi, Z_t, a_t^\pi) + r_T(p_T^\pi, Z_T)\right)$$

with control costs:

- Rewards at $t = 0, \ldots, T-1$ from decision a in state (p, z)

$$r_t : P \times \mathbb{R}^d \times A \to \mathbb{R} \qquad (p, z, a) \mapsto r_t(p, z, a)$$

- Scrap value at $t = T$, no action:

$$r_T : P \times \mathbb{R}^d \to \mathbb{R} \qquad (p, z) \mapsto r_T(p, z)$$

Target

Determine a policy $\pi^* = (\pi_t^*)_{t=0}^{T-1}$ which maximizes

$$\pi \mapsto v_0^\pi(p_0, z_0) = \mathbb{E}\left(\sum_{t=0}^{T-1} r_t(p_t^\pi, Z_t, a_t^\pi) + r_T(p_T^\pi, Z_T)\right)$$

over all policies.

Any maximizer is called optimal policy, and is denoted by

$$\pi^* = (\pi_t^*)_{t=0}^{T-1}$$

Example: Bermudan Put option

with strike K, at interest rate $\rho \geq 0$, for maturity T has fair price

$$\sup_{\tau}\{\mathbb{E}(e^{-\rho\tau}(K - Z_{\tau})^{+}, 0))$$

over all $\{0, 1, \ldots, T\}$-valued stopping times τ.

Continuous part uncontrolled: $(Z_t)_{t=0}^{T}$ follows

$$Z_{t+1} = W_{t+1} Z_t, \qquad Z_0 = z_0 \in]0, \infty[$$

where $(W_t)_{t=1}^{T}$ are iid log-normal variables.

Example: Bermudan Put option

Discrete part:

- Positions $P = \{\text{stopped}, \text{goes}\}$
- Actions $A = \{\text{stop}, \text{go}\}$
- Position change

$$\begin{bmatrix} \alpha(\text{stopped}, \text{stop}) & \alpha(\text{stopped}, \text{go}) \\ \alpha(\text{goes}, \text{stop}) & \alpha(\text{goes}, \text{go}) \end{bmatrix} = \begin{bmatrix} \text{stopped} & \text{stopped} \\ \text{stopped} & \text{goes} \end{bmatrix}.$$

Thus we have with $P = \{1, 2\}$, and $A = \{1, 2\}$.

$$(\alpha(p, a))_{p,a=1}^{2} \sim \begin{bmatrix} \alpha(1, 1) & \alpha(2, 1) \\ \alpha(1, 2) & \alpha(2, 2) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix},$$

Example: Bermudan Put option

The reward at time $t = 0, \ldots, T-1$ and scrap value are

$$\begin{aligned} r_t(p,z,a) &= e^{-\rho t}(K-z)^{+}(p-\alpha(p,a)), \\ r_T(p,z) &= e^{-\rho T}(K-z)^{+}(p-\alpha(p,1)), \end{aligned}$$

for $p \in P$, $a \in A$, $z \in \mathbb{R}_+$

Theoretical solution

Define the original Bellman operator

$$\mathcal{T}_t v(p,z) = \max_{a\in A}\left(r_t(p,z,a) + \sum_{p'\in P}\alpha^a_{p,p'}\mathbb{E}(v(p',W_{t+1}z))\right),$$

and introducer the *Bellman recursion* (backward induction)

$$v_T = r_T, \quad v_t = \mathcal{T}_t v_{t+1} \qquad \text{for } t = T-1,\dots,0.$$

There exists a recursive solution $(v^*_t)^T_{t=0}$, called *value functions*, they determines an optimal policy $\pi^* = (\pi_t)^{T-1}_{t=0}$ via

$$\pi^*_t(p,z) = \mathrm{argmax}_{a\in A}\left(r_t(p,z,a) + \sum_{p'\in P}\alpha^a_{p,p'}\mathbb{E}(v^*_{t+1}(p',W_{t+1}z)\right)$$

for all $p\in P, z\in\mathbb{R}^d, t = 0,\dots,T-1$.

Numerical solution

If reward and scrap functions are convex, then

instead of the original Bellman operator

$$\mathcal{T}_t v(p, z) = \max_{a \in A} \left(r_t(p, z, a) + \sum_{p' \in P} \alpha^a_{p,p'} \mathbb{E}(v(p', W_{t+1} z)) \right),$$

we consider the modified Bellman operator

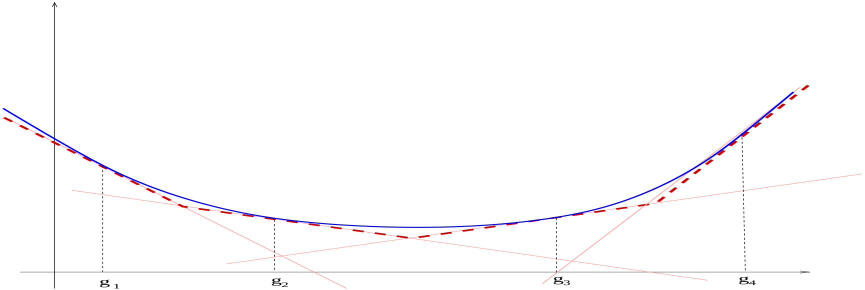
$$\mathcal{T}_t^{m,n}(p, .) = \mathcal{S}_{G^m} \max_{a \in A} \left(r_t(p, \cdot, a) + \sum_{p' \in P} \alpha^a_{p,p'} \sum_{k=1}^{n} \nu_{t+1}(k) v(p', W_{t+1}(k) \cdot) \right)$$

For convex $v(p,\cdot)$,

the modified Bellman operator is

$$\mathcal{T}_t^{m,n}(p,.) = \mathcal{S}_{G^m} \max_{a \in A} \left(r_t(p,\cdot,a) + \sum_{p' \in P} \alpha^a_{p,p'} \sum_{k=1}^{n} \nu_{t+1}(k) v(p', W_{t+1}(k)\cdot) \right)$$

where $\mathcal{S}_{G^m}$ stands for the sub-gradient envelope for the grid $G^m = \{g^1, \ldots, g^m\}$:



Modified backward induction

Using modified Bellman operators $\mathcal{T}^{m,n}$, we introduce backward induction

$$\begin{aligned} v_T^{m,n} &= \mathcal{S}_{G^m} r_T\,, \\ v_t^{m,n} &= \mathcal{T}_t^{m,n} v_{t+1}^{m,n}, \qquad t = T-1, \ldots 0. \end{aligned}$$

which enjoys excellent asymptotic properties.

Using matrix representations of convex piecewise linear functions, the modified backward induction boils down to simple linear algebra.

Using further approximations and techniques from data mining (hierarchical clustering, next neighbor search) we obtain very efficient implementations

Main problem

How far is an approximate solution is from the optimal one?
For optimal stopping: Duality idea of C. Rogers

Upper bound estimation: Let $(Z_t)_{t=0}^T$ be adapted, and $\mathcal{V}$ be all finite stopping times.

The optimal stopping value is attained at some stopping time τ^*

$$V_0^* := \sup_{\tau \in \mathcal{V}} \mathbb{E}(Z_\tau) = \mathbb{E}(Z_{\tau^*})$$

and dominated by the expectation of a pathwise maximum

$$V_0^* := \sup_{\tau \in \mathcal{V}} \mathbb{E}(Z_\tau) \leq \mathbb{E}(\sup_{0 \leq t \leq T} Z_t).$$

Duality idea of C. Rogers:

Subtracting any martingale $(M_t)_{t=0}^T \in \mathcal{M}_0$ starting at the origin $M_0 = 0$, we have

$$V_0^* = \sup_{\tau \in \mathcal{V}} \mathbb{E}(Z_\tau - M_\tau) \leq \mathbb{E}(\sup_{0 \leq t \leq T} (Z_t - M_t)).$$

this estimate is tight and is attained at some martingale $(M_t^*)_{t=0}^T$

$$V_0^* = \mathbb{E}(\sup_{0 \leq t \leq T} (Z_t - M_t^*)).$$

Duality idea of C. Rogers

Random upper bound: Given simulated sample paths of $(Z_t - M_t)_{t=0}^T$, determine the maximum on each trajectory and calculate their empirical mean.

There are many ideas how to chose the best martingale (close to $(M_t^*)_{t=0}^T$)

Random lower bound: Take some stopping time τ, stop trajectories of $(Z_t - M_t)_{t=0}^T$ and average.

Self-tuning: The closer the stopping time τ and the martingale $(M_t)_{t=0}^T$ are to their optimal counterparts τ^* and $(M_t^*)_{t=0}^T$, the narrower the bounds, the lower Monte-Carlo variance.

Bound estimation

for our stochastic switching systems, the arguments are similar, but instead of martingale we have a family of martingale increments.

Main problem

Given: A numerical scheme returns approximate value functions $(v_t)_{t=0}^T$, approximate expected value functions $(v_t^E)_{t=0}^T$ along with corresponding policy $(\pi_t)_{t=0}^{T-1}$ given by

$$\pi_t(p,z) = \operatorname{argmax}(r_t(p,z,a) + \sum_{p' \in P} \alpha_{p,p'}^a v_{t+1}^E(p',z)))$$

Question: How far we are from the optimality? In other words, at a given a point (p_0, z_0), estimate the performance gap

$$[v_0^{\pi}(p_0,z_0), v_0^{\pi^*}(p_0,z_0)].$$

Solution by bounds estimation:

Explicit construction of random variables

$$\underline{v}_0^{\pi,\varphi}(p_0, z_0), \quad \overline{v}_0^{\varphi}(p_0, z_0)$$

satisfying

$$\mathbb{E}(\underline{v}_0^{\pi,\varphi}(p_0, z_0)) = v_0^{\pi}(p_0, z_0) \leq v_0^{\pi^*}(p_0, z_0) \leq \mathbb{E}(\bar{v}_0^{\varphi}(p_0, z_0)).$$

Using MC, one estimates both means with confidence bounds to understand the performance gap.

Self-tuning: The better the approximate solution $(v_t)_{t=0}^T$ $(v_t^E)_{t=0}^T$, the narrower the gap, the lower the variance of MC.

We prove inductively

Lower bound (variance reduction)

1) Given approximate solution $(v_t)_{t=0}^T$ $(v_t^E)_{t=0}^T$ with the corresponding policy $(\pi_t)_{t=0}^{T-1}$, implement control variables $(\varphi_t)_{t=1}^T$ as

$$\varphi_t(p,z,a) = \sum_{p' \in P} \alpha_{p,p'}^a (\frac{1}{I} \sum_{i=1}^I v_t(p', W_t^{(i)} z) - v_t(p', W_t z)),$$

for all $p \in P$, $a \in A$, $z \in \mathbb{R}^d$, where $(W_t^{(1)}, \dots, W_t^{(I)}, W_t)$ are independent identically distributed.

2) Chose a number $K \in \mathbb{N}$ of Monte-Carlo trials and obtain for $k = 1, \dots, K$ independent realizations $(W_t(\omega_k))_{t=1}^T$ of disturbances.

3) Starting at $z_0^k := z_0 \in \mathbb{R}^d$, define for $k = 1, \dots, K$ trajectories $(z_t^k)_{t=0}^T$ recursively

$$z_{t+1}^k = W_{t+1}(\omega_k) z_t^k, \qquad t = 0, \dots, T-1$$

and determine realizations

$$\varphi_t(p, z_{t-1}^k, a)(\omega_k), \qquad t = 1, \dots, T, \quad k = 1, \dots, K.$$

4) For each $k = 1, \ldots, K$ initialize the recursion at $t = T$ as

$$\underline{v}_T^{\pi,\varphi}(p, z_T^k)(\omega_k) = r_T(p, z_T^k) \qquad \text{for all } p \in P$$

and continue for $t = T-1, \ldots, 0$ and for all $p \in P$ by

$$\begin{aligned} \underline{v}_t^{\pi,\varphi}(p, z_t^k)(\omega_k) = r_t(p, z_t^k, \pi_t(p, z_t^k)) + \varphi_{t+1}(p, z_t^k, \pi_t(p, z_t^k))(\omega_k) \\ + \sum_{p' \in P} \alpha_{p,p'}^{\pi_t(p, z_t^k)} \underline{v}_{t+1}^{\pi,\varphi}(p', z_{t+1}^k)(\omega_k) \end{aligned}$$

5) Calculate sample mean

$$\frac{1}{K} \sum_{k=1}^{K} v_0^{\pi,\varphi}(p_0, z_0)(\omega_k)$$

to estimate $\mathbb{E}(\underline{v}_0^{\pi,\varphi}(p_0, z_0))$ with confidence bounds.

In multi-period case we prove inductively

Upper bound (duality of C. Rogers)

replace in the step 4)

$$\underline{v}_t^{\pi,\varphi}(p, z_t^k)(\omega_k) = r_t(p, z_t^k, \pi_t(p, z_t^k)) + \varphi_{t+1}(p, z_t^k, \pi_t(p, z_t^k))(\omega_k) + \sum_{p' \in P} \alpha_{p,p'}^{\pi_t(p,z_t^k)} \underline{v}_{t+1}^{\pi,\varphi}(p', z_{t+1}^k)(\omega_k)$$

by

$$\overline{v}_t^{\varphi}(p, z_t^k)(\omega_k) = \max_{a \in A} \Big(r_t(p, z_t^k, a) + \varphi_{t+1}(p, z_t^k, a)(\omega_k) + \sum_{p' \in P} \alpha_{p,p'}^{a} \overline{v}_{t+1}^{\varphi}(p', z_{t+1}^k)(\omega_k) \Big)$$

with the same initialization

$$\overline{v}_T^{\varphi}(p, z_T^k)(\omega_k) = r_T(p, z_T^k) \qquad \text{for all } p \in P$$

Illustration Bermudan Put

S_0	σ	maturity	confidence interval	LSM mean	LSM se
36	0.2	1	[4.4763, 4.4768]	4.472	.0100
36	0.2	2	[4.8296, 4.8312]	4.821	.0120
36	0.4	1	[7.0989, 7.0992]	7.091	.0200
36	0.4	2	[8.4965, 8.4968]	8.488	.0240
38	0.2	1	[3.2481, 3.2489]	3.244	.0090
38	0.2	2	[3.7355, 3.7370]	3.735	.0110
38	0.4	1	[6.1451, 6.1452]	6.139	.0190
38	0.4	2	[7.6580, 7.6583]	7.669	.0220
40	0.2	1	[2.3119, 2.3129]	2.313	.0090
40	0.2	2	[2.8765, 2.8776]	2.879	.0100
40	0.4	1	[5.3093, 5.3094]	5.308	.0180

Illustration Bermudan Put

S_0	σ	maturity	confidence interval	LSM mean	LSM se
40	0.4	1	[5.3093, 5.3094]	5.308	.0180
40	0.4	2	[6.9075, 6.9077]	6.921	.0220
42	0.2	1	[1.6150, 1.6158]	1.617	.0070
42	0.2	2	[2.2053, 2.2060]	2.206	.0100
42	0.4	1	[4.5797, 4.5798]	4.588	.0170
42	0.4	2	[6.2351, 6.2354]	6.243	.0210
44	0.2	1	[1.1081, 1.1087]	1.118	.0070
44	0.2	2	[1.6836, 1.6843]	1.675	.0090
44	0.4	1	[3.9449, 3.9450]	3.957	.0170
44	0.4	2	[5.6324, 5.6326]	5.622	.0210

Swing option numerical results

	CSS	MH
Position (Rights + 1)	confidence interval	confidence interval
2	[4.737, 4.761]	[4.773, 4.794]
3	[9.005, 9.031]	[9.016, 9.091]
4	[13.001, 13.026]	[12.959, 13.100]
5	[16.805, 16.830]	[16.773, 16.906]
6	[20.465, 20.491]	[20.439, 20.580]
11	[37.339, 37.363]	[37.305, 37.540]
16	[52.694, 52.718]	[52.670, 53.009]
21	[67.070, 67.095]	[67.050, 67.525]
31	[93.811, 93.835]	[93.662, 94.519]

Swing option numerical results

	CSS	MH
Position (Rights + 1)	confidence interval	confidence interval
41	[118.639, 118.663]	[118.353, 119.625]
51	[142.059, 142.084]	[141.703, 143.360]
61	[164.368, 164.392]	[163.960, 166.037]
71	[185.757, 185.781]	[185.335, 187.729]
81	[206.362, 206.386]	[205.844, 208.702]
91	[226.284, 226.308]	[225.676, 228.985]
101	[245.601, 245.625]	[244.910, 248.651]

Asset liquidation: Position control

Remember: $p \in P$ is the number of asset units. Actions are

$$A = \{0, \ldots, a_{\text{max}}\} \times \{1\} \cup \{0, \ldots, a_{max}\} \times \{2\}$$

with the interpretation that $(a, 1)$, $(a, 2)$ stand for the limit and market order of size $a = 0, \ldots, a_{\text{max}}$ respectively.

Asset liquidation: Position control

For illustration, we use

1. Limit orders: $\alpha^{(a,1)}_{p,(p-a)\vee 0} = \begin{cases} 0.3 & \text{if } a = 1; \\ 0.2 & \text{if } a = 2; \\ 0.1 & \text{if } a = 3; \end{cases}$ and $\alpha^{(a,1)}_{p,p} = 1 - \alpha^{(a,1)}_{p,(p-a)\vee 0}$.
2. Market orders: $\alpha^{(a,2)}_{p,(p-a)\vee 0} = \begin{cases} 1 & \text{if } a = 1; \\ 0.9 & \text{if } a = 2; \\ 0.8 & \text{if } a = 3; \end{cases}$ and $\alpha^{(a,2)}_{p,p} = 1 - \alpha^{(a,2)}_{p,(p-a)\vee 0}$.

Asset liquidation: Spread evolution

We model it as auto-regression and realize as the first component $(Z_t^{(1)})_{t\in\mathbb{N}}$ of the linear state space process $(Z_t)_{t\in\mathbb{N}}$ defined by the recursion

$$\underbrace{\begin{bmatrix} Z_{t+1}^{(1)} \\ Z_{t+1}^{(2)} \end{bmatrix}}_{Z_{t+1}} = \underbrace{\begin{bmatrix} -\phi & \sigma N_{t+1} \\ 0 & 1 \end{bmatrix}}_{W_{t+1}} \underbrace{\begin{bmatrix} Z_t^{(1)} \\ Z_t^{(2)} \end{bmatrix}}_{Z_t}, \qquad \begin{bmatrix} Z_0^{(1)} \\ Z^{(2)} \end{bmatrix} = \begin{bmatrix} z_0 \\ 1 \end{bmatrix}$$

where $(N_t)_{t\in\mathbb{N}}$ is an iid sequence.

Asset liquidation: Reward functions

are given by

$$\begin{aligned} r_t(p,z,a) &= -g_t(b_t - p) - (\mu + z^{(1)})(a_2 - 1), \quad t = 0, \ldots, T-1 \\ r_T(p,z) &= -g_T(b_T - p). \end{aligned}$$

where $-(\mu + z^{(1)})(a_2 - 1)$ is a loss from crossing the spread when placing market order $a_2 = 2$

where $g_t(b_t - p)$ is a penalty on the deviation $b_t - p$ of the current long position p from a pre-determined benchmark level $b_t \in \mathbb{R}$.

we use different time-dependent penalizations

		CSS		Lower Bound		Upper Bound	
γ_t	z_0	Point	Range	Point	Range	Point	Range
1	-1	-26.9405	-6.0797	-26.8207(.0124)	-5.9593(.0124)	-26.8203(.0124)	-5.9579(.0121)
	0	-28.7044	-7.8424	-28.5869(.0121)	-7.7243(.0121)	-28.5865(.0121)	-7.7231(.0118)
	1	-27.4605	-6.6270	-27.3505(.0128)	-6.5169(.0128)	-27.3500(.0128)	-6.5158(.0125)
$\frac{1}{50}t$	-1	-15.7198	-4.2042	-15.5985(.0124)	-4.0832(.0124)	-15.5971(.0124)	-4.0822(.0120)
	0	-17.0048	-5.5145	-16.8850(.0121)	-5.3955(.0121)	-16.8833(.0121)	-5.3944(.0117)
	1	-15.9406	-4.4212	-15.8275(.0127)	-4.3092(.0127)	-15.8260(.0127)	-4.3082(.0124)
$\frac{1}{20}t$	-1	-29.4207	-6.3232	-29.3147(.0110)	-6.2028(.0110)	-29.3140(.0110)	-6.2017(.0123)
	0	-30.8077	-7.7448	-30.7045(.0105)	-7.6270(.0105)	-30.7037(.0105)	-7.6260(.0120)
	1	-29.6721	-6.5819	-29.5759(.0112)	-6.4709(.0112)	-29.5752(.0112)	-6.4698(.0127)

Conclusion

for switching problems,

- there is similarity to optimal stopping since stochastic dynamics is uncontrolled
- an adaptation of duality estimates is possible
- instead of martingale, we have a family of martingale increments
- we provide a unified view on variance reduction and duality
- we suggest constructing martingale increments from approximate solution
- we obtain tight bounds for practical problems

Thank you!

Round table: Stochastic control

- algorithmic approach
- high-dimensional stochastic control
- duality & pathwise diagnostics
- applications

Algorithmic approach / high-dimensional control

- focus on narrow class (controlled linear dynamics)
- apply high-dimensional regression with localization
- statistical tools (hierarchical clustering, next neighbor search)

Duality & pathwise techniques

- extend duality to controlled linear dynamics (reference measure)
- asymptotic analysis for reference measures
- numerical studies

Applications

- Real options
- Market equilibria
- Artificial intelligence