Exponential Ergodicity in a Sobolev Space

Applications to Reinforcement Learning, and ...

A. M. Devraj, I. Kontoyiannis and S. Meyn

Laboratory for Cognition & Control in Complex Systems
University of Florida, Gainesville

PDE and Probability Methods for Interactions

March 30-31, 2017

Thanks to NSF

Outline

Goals

Markov process $\boldsymbol{X}$ on state space $\mathsf{X} = \mathbb{R}^\ell$
Transition semigroup: for $t \geq 0$, $x \in \mathsf{X}$, $A \in \mathcal{B}$,

$$P^t(x, A) := \mathsf{P}_x\{X(t) \in A\} := \Pr\{X(t) \in A \mid X(0) = x\}$$

Operator notation: for $f : \mathsf{X} \to \mathbb{R}$, signed measure ν on $(\mathsf{X}, \mathcal{B})$:

$$P^t f\,(x) = \int f(y) P^t(x, dy)$$

$$\nu P^t\,(A) = \int \nu(dx) P^t(x, A)$$

Generator $\mathcal{D}$ $\quad$ ($= P - I$ in discrete time)

Goals

Markov process on $\mathsf{X} = \mathbb{R}^\ell$ (continuous or discrete time)
Generator $\mathcal{D}$ $\quad (= P - I$ in discrete time)

Goals

Markov process on $\mathsf{X} = \mathbb{R}^\ell$ (continuous or discrete time)
Generator $\mathcal{D}$ $\quad (= P - I$ in discrete time)

Compute or bound solution $h \colon \mathsf{X} \to \mathbb{C}$

1. Eigenfunction: $\mathcal{D}h = \lambda h$

Goals

Markov process on $\mathsf{X} = \mathbb{R}^\ell$ (continuous or discrete time)
Generator $\mathcal{D}$ $\quad (= P - I$ in discrete time)

Compute or bound solution $h\colon \mathsf{X} \to \mathbb{C}$

1. Eigenfunction: $\mathcal{D}h = \lambda h$
2. Poisson eqn: $\mathcal{D}h = -\tilde{c}$ [appl. to simulation and control]
3. Dirichlet+: $\mathcal{D}h = \mathcal{G}(h, \nabla h)$ [appl. to simulation and control]

Goals

Markov process on $\mathsf{X} = \mathbb{R}^\ell$ (continuous or discrete time)
Generator $\mathcal{D}$ $\quad (= P - I$ in discrete time)

Compute or bound solution $h \colon \mathsf{X} \to \mathbb{C}$

1. Eigenfunction: $\mathcal{D}h = \lambda h$
2. Poisson eqn: $\mathcal{D}h = -\tilde{c}$ [appl. to simulation and control]
3. Dirichlet+: $\mathcal{D}h = \mathcal{G}(h, \nabla h)$ [appl. to simulation and control]
4. Its gradient $K = \nabla h$ [appl. to nonlinear filtering – Mehta & M.]

Goals

Markov process on $\mathsf{X} = \mathbb{R}^\ell$ (continuous or discrete time)
Generator $\mathcal{D}$ $\quad (= P - I$ in discrete time)

Compute or bound solution $h \colon \mathsf{X} \to \mathbb{C}$

1. Eigenfunction: $\mathcal{D}h = \lambda h$
2. Poisson eqn: $\mathcal{D}h = -\tilde{c}$ [appl. to simulation and control]
3. Dirichlet+: $\mathcal{D}h = \mathcal{G}(h, \nabla h)$ [appl. to simulation and control]
4. Its gradient $K = \nabla h$ [appl. to nonlinear filtering – Mehta & M.]

Approach: New operator norm for spectral theory.

Stick to discrete time here

Notation and Assumptions

Operator notation for geometric ergodicity (M&T and K&M):
$v\colon \mathsf{X} \to [1, \infty)$ continuous "weighting function".
For $f\colon \mathsf{X} \to \mathbb{R}$,

$$\|f\|_v := \sup_x \frac{|f(x)|}{v(x)}.$$

Corresponding Banach space:

$$L_\infty^v := \{f\colon \mathsf{X} \to \mathbb{R} : \|f\|_v < \infty\}$$

Notation and Assumptions

Operator notation for geometric ergodicity (M&T and K&M):
$v\colon \mathsf{X} \to [1,\infty)$ continuous "weighting function".
For $f\colon \mathsf{X} \to \mathbb{R}$,

$$\|f\|_v := \sup_x \frac{|f(x)|}{v(x)}.$$

Corresponding Banach space:

$$L_\infty^v := \{f\colon \mathsf{X} \to \mathbb{R} : \|f\|_v < \infty\}$$

Geometric ergodicity: There is $b_0 < \infty$, $\varrho_0 < 1$ such that

$$|\!|\!|\widetilde{P}^t|\!|\!|_v \le b_0 \rho_0^t\,, \quad t \ge 0; \qquad \widetilde{P}^t = P^t - \mathbf{1} \otimes \pi\,.$$

$\Longleftrightarrow$ spectral gap in L_∞^v

Notation and Assumptions

$\|f\|_v := \sup_x |f(x)|/v(x)$

$$\text{New norm:} \quad \|f\|_{v,k} = \max_{|\alpha|\le k} \|\partial^\alpha f\|_v, \qquad k \ge 1$$

Notation and Assumptions

$\|f\|_v := \sup_x |f(x)|/v(x)$

New norm: $\|f\|_{v,k} = \max_{|\alpha|\leq k} \|\partial^\alpha f\|_v, \qquad k \geq 1$

New Banach spaces:

$$L_\infty^{v,0} = \{f \in L_\infty^v : f \text{ is continuous}\}$$
$$L_\infty^{v,k} = \{g\colon \mathsf{X} \to \mathbb{R} : \partial^\alpha f \in L_\infty^{v,0} \text{ for all } |\alpha| \leq k\}$$

Induced operator norm, for any kernel $\widehat{P}$,

$$\|\widehat{P}\|_{v,k} := \sup\left\{\frac{\|\widehat{P}h\|_{v,k}}{\|h\|_{v,k}} : h \in L_\infty^{v,k},\ \|h\|_{v,k} \neq 0\right\}.$$

Notation and Assumptions

$\|f\|_v := \sup_x |f(x)|/v(x)$

$$\text{New norm:} \quad \|f\|_{v,k} = \max_{|\alpha| \le k} \|\partial^\alpha f\|_v, \qquad k \ge 1$$

New Banach spaces:

$$L_\infty^{v,0} = \{f \in L_\infty^v : f \text{ is continuous}\}$$
$$L_\infty^{v,k} = \{g \colon \mathsf{X} \to \mathbb{R} : \partial^\alpha f \in L_\infty^{v,0} \text{ for all } |\alpha| \le k\}$$

Induced operator norm, for any kernel $\widehat{P}$,

$$|\!|\!|\widehat{P}|\!|\!|_{v,k} := \sup\Big\{ \frac{\|\widehat{P}h\|_{v,k}}{\|h\|_{v,k}} : h \in L_\infty^{v,k},\ \|h\|_{v,k} \neq 0 \Big\}.$$

We will stick to $k = 1$:

$$\|f\|_{v,1} = \max\Big\{ \|f\|_v, \|\partial^1 f\|_v, \dots, \|\partial^\ell f\|_v \Big\}$$

Notation and Assumptions

Markovian system dynamics

$$X(t+1) = a(X(t), N(t+1)), \quad t \in \mathbb{Z}_+, \boldsymbol{N} \text{ i.i.d.}$$
$$P(x, A) = \mathsf{P}(a(x, N(1)) \in A)$$

Notation and Assumptions

Markovian system dynamics

$$X(t+1) = a(X(t), N(t+1)), \quad t \in \mathbb{Z}_+, \boldsymbol{N} \text{ i.i.d.}$$
$$P(x, A) = \mathsf{P}(a(x, N(1)) \in A)$$

A1 Smooth dynamics: $a : \mathbb{R}^{d \times m} \to \mathbb{R}^d$ is C^1 and Lipschitz in x

A2 Densities: For some $t_0 \geq 1$ and C^1 function p_{t_0},

$$P^{t_0}(x, A) = \int_A p_{t_0}(x, y)\, dy\,, \qquad x \in \mathsf{X},\ A \in \mathcal{B}$$

A3 ψ-irreducibility: for some $x_0 \in \mathsf{X}$,

$$P^t(x, O) > 0 \qquad \text{all } t = t_x \geq 0 \text{ sufficiently large, each nbd } O \text{ of } x_0$$

Notation and Assumptions

Markovian system dynamics

$$X(t+1) = a(X(t), N(t+1)), \quad t \in \mathbb{Z}_+, \boldsymbol{N} \text{ i.i.d.}$$
$$P(x, A) = \mathsf{P}(a(x, N(1)) \in A)$$

A1 Smooth dynamics: $a : \mathbb{R}^{d\times m} \to \mathbb{R}^d$ is C^1 and Lipschitz in x

A2 Densities: For some $t_0 \geq 1$ and C^1 function p_{t_0},

$$P^{t_0}(x, A) = \int_A p_{t_0}(x, y)\, dy\,, \qquad x \in \mathsf{X},\ A \in \mathcal{B}$$

A3 ψ-irreducibility: for some $x_0 \in \mathsf{X}$,

$$P^t(x, O) > 0 \qquad \text{all } t = t_x \geq 0 \text{ sufficiently large, each nbd } O \text{ of } x_0$$

A4 Donsker-Varadhan drift condition:

$$\mathcal{H}(V) := \log(Pe^V) - V \leq -\delta W + b\mathbb{I}_C\,, \qquad \text{(DV3)}$$

$V = \log(v)$ and $W \in L_\infty^{v,0}$ coercive, C compact. [K&M, L. Wu]

Main Result: Separability in $L_\infty^{v,1}$

A1 Smooth dynamics

A2 Densities

A3 ψ-irreducibility

A4 Donsker-Varadhan drift condition:

$$\mathcal{H}(V) \le -\delta W + b\mathbb{I}_C\,. \tag{DV3}$$

Theorem: *Separability in $L_\infty^{v,1}$*

The kernel P^t is *separable* in $L_\infty^{v,1}$ for some t_1 and all $t \ge t_1$:

$$P^t \approx \sum_{k=1}^{n} s_k \otimes \nu_k\,, \qquad \textit{approximation in } L_\infty^{v,1}$$

Main Result: Separability in $L_\infty^{v,1}$

Theorem: *Separability in* $L_\infty^{v,1}$

The kernel P^t is *separable* in $L_\infty^{v,1}$ for some t_1 and all $t \geq t_1$:

$$P^t \approx \sum_{k=1}^{n} s_k \otimes \nu_k \,, \qquad \textit{approximation in } L_\infty^{v,1}$$

Corollaries: Discrete spectrum and ... there is $b_0 < \infty$ and $\varrho_0 < 1$ s.t.

$$\|\widetilde{P}^t\|_{v,1} \leq b_0 \varrho_0^t, \quad t \geq t_1.$$

Main Result: Separability in $L_\infty^{v,1}$

Theorem: *Separability in $L_\infty^{v,1}$*

The kernel P^t is *separable* in $L_\infty^{v,1}$ for some t_1 and all $t \geq t_1$:

$$P^t \approx \sum_{k=1}^{n} s_k \otimes \nu_k \,, \qquad \textit{approximation in } L_\infty^{v,1}$$

Corollaries: Discrete spectrum and ... there is $b_0 < \infty$ and $\varrho_0 < 1$ s.t.

$$\|\widetilde{P}^t\|_{v,1} \leq b_0 \varrho_0^t, \quad t \geq t_1.$$

Interpretation: for $f \in L_\infty^{v,1}$, $\quad P^t f\,(x) \to \pi(f)\,, \quad \partial^i P^t f\,(x) \to 0,$

uniform geometric convergence rate.

Proof: Truncation of P^{t_1} to compact domain, as in [K&M 200X];
$\text{Spectrum}_{L_\infty^{v,1}}(P^t) \subseteq \text{Spectrum}_{L_\infty^{v}}(P^t)$

Connection with Lyapunov Exponents

Remains a mystery

Sensitivity process: $\mathcal{S}^{\intercal}(t) = \dfrac{\partial}{\partial X(0)} X(t)$

$$\mathcal{S}(t+1) = \mathcal{A}(t+1)\mathcal{S}(t), \quad \mathcal{S}(0) = I$$

where $\mathcal{A}^{\intercal}(t) := \nabla_x a\left(X(t-1), N(t)\right)$.

Connection with Lyapunov Exponents

Remains a mystery

Sensitivity process: $\mathcal{S}^{\intercal}(t) = \dfrac{\partial}{\partial X(0)} X(t)$

$$\mathcal{S}(t+1) = \mathcal{A}(t+1)\mathcal{S}(t), \quad \mathcal{S}(0) = I$$

where $\mathcal{A}^{\intercal}(t) := \nabla_x a\,(X(t-1), N(t))$.

Lyapunov exponents: $\Lambda = \lim_{t\to\infty} \dfrac{1}{t} \log \|\mathcal{S}(t)\| \quad a.s.$

$$\Lambda_p = \lim_{t\to\infty} \frac{1}{t} \log \mathsf{E}[\|\mathcal{S}(t)\|^p]$$

Gradient representation:

$$\boxed{\nabla P^t = Q^t \nabla}$$

$$Q^t g(x) := \mathsf{E}_x\left[\mathcal{S}^{\intercal}(t) g(X(t))\right], \quad g = \nabla c.$$

Discounted cost value function

Cost function: $c : \mathbb{R}^d \to \mathbb{R}$

Discount factor: $\alpha < 1$

Value function: $$h_\alpha(x) = \sum_{t=0}^{\infty} \alpha^t \mathsf{E}[c(X(t)) \mid X(0) = x]$$

Discounted cost value function

Cost function: $c : \mathbb{R}^d \to \mathbb{R}$
Discount factor: $\alpha < 1$
Value function: $$h_\alpha(x) = \sum_{t=0}^{\infty} \alpha^t \mathsf{E}[c(X(t)) \mid X(0) = x]$$

Goal of TD learning: Approximate h_α in parameterized class $\{h_\alpha^\theta : \theta \in \mathbb{R}^\ell\}$

Discounted cost value function

Cost function: $c : \mathbb{R}^d \to \mathbb{R}$
Discount factor: $\alpha < 1$
Value function: $$h_\alpha(x) = \sum_{t=0}^{\infty} \alpha^t \mathsf{E}[c(X(t)) \mid X(0) = x]$$

Goal of TD learning: Approximate h_α in parameterized class $\{h_\alpha^\theta : \theta \in \mathbb{R}^\ell\}$
Goal of ∇-TD learning: (new)

$$\theta^* = \arg\min_\theta \mathsf{E}[\|\nabla h_\alpha^\theta(X) - \nabla h_\alpha(X)\|^2], \quad X \sim \pi$$

Discounted cost value function

Cost function: $c : \mathbb{R}^d \to \mathbb{R}$
Discount factor: $\alpha < 1$
Value function: $$h_\alpha(x) = \sum_{t=0}^{\infty} \alpha^t \mathsf{E}[c(X(t)) \mid X(0) = x]$$

Goal of TD learning: Approximate h_α in parameterized class $\{h_\alpha^\theta : \theta \in \mathbb{R}^\ell\}$
Goal of ∇-TD learning: (new)

$$\theta^* = \arg\min_\theta \mathsf{E}[\|\nabla h_\alpha^\theta(X) - \nabla h_\alpha(X)\|^2] \,, \quad X \sim \pi$$

Example: affine parameterization,

$$h_\alpha^\theta(x) = \kappa(\theta) + \sum_{j=1}^{\ell} \theta_j \psi_j(x), \qquad \psi : \mathbb{R}^d \to \mathbb{R}^\ell$$

Discounted cost value function

Value function: $$h_\alpha(x) = \sum_{t=0}^{\infty} \alpha^t \mathsf{E}[c(X(t)) \mid X(0) = x]$$

Goal of ∇-TD learning: $$\theta^* = \arg\min_\theta \mathsf{E}_\pi[\|\nabla h_\alpha^\theta(X) - \nabla h_\alpha(X)\|^2]$$

Recover missing constant:

$$h_\alpha^\theta(x) = \theta^{\intercal}\psi(x) + \kappa(\theta)$$

$$\kappa(\theta) = -\theta^{\intercal}\pi(\psi) + \pi(c)/(1-\alpha)$$

$$\Longrightarrow \quad \pi(h_\alpha^\theta) = \pi(h_\alpha) = \pi(c)/(1-\alpha)$$

Gradient Representation

Goal of ∇-TD learning:

$$\theta^* = \arg\min_\theta \mathsf{E}[\|\nabla h_\alpha^\theta(X) - \nabla h_\alpha(X)\|^2], \quad X \sim \pi$$

Representation:

$$\nabla h_\alpha(x) = \sum_{t=0}^{\infty} \alpha^t \nabla P^t c(x)$$

$$\Longrightarrow \quad \nabla h_\alpha = \Omega_\alpha \nabla c := \sum_{t=0}^{\infty} \alpha^t Q^t \nabla c(x)$$

$$= \sum_{t=0}^{\infty} \alpha^t \mathsf{E}\big[\mathcal{S}^{\mathsf{T}}(t) \nabla c(X(t)) \mid X(0) = x\big]$$

Adjoint Representation of θ^*

Goal of ∇-TD learning:

$$\theta^* = \arg\min_\theta \mathsf{E}[\|\nabla h_\alpha^\theta(X) - \nabla h_\alpha(X)\|^2], \quad X \sim \pi$$

Solution:

$$\theta^* = M^{-1} b$$

$$M = \mathsf{E}_\pi\Big[(\nabla\psi(X))^{\mathsf{T}} \nabla\psi(X)\Big]$$
$$b = \mathsf{E}_\pi\Big[(\nabla\psi(X))^{\mathsf{T}} \nabla h_\alpha(X)\Big]$$

$$\nabla h_\alpha = \Omega_\alpha \nabla c$$

Adjoint Representation of θ^*

Goal of ∇-TD learning:

$$\theta^* = \arg\min_{\theta} \mathsf{E}[\|\nabla h_\alpha^\theta(X) - \nabla h_\alpha(X)\|^2], \quad X \sim \pi$$

Solution:

$$\theta^* = M^{-1} b$$

$$\begin{aligned} M &= \mathsf{E}_\pi\left[(\nabla\psi(X))^{\intercal}\nabla\psi(X)\right] \\ b &= \mathsf{E}_\pi\left[(\nabla\psi(X))^{\intercal}\nabla h_\alpha(X)\right] \end{aligned}$$

$$\nabla h_\alpha = \Omega_\alpha \nabla c$$

Adjoint gives causal representation:

$$\begin{aligned} b_i &= \langle \partial^i \psi, \Omega_\alpha \nabla c \rangle \\ &= \langle \Omega_\alpha^\dagger \partial^i \psi, \nabla c \rangle \end{aligned}$$

Adjoint Representation of θ^*

Goal of ∇-TD learning:

$$\theta^* = \arg\min_{\theta} \mathsf{E}[\|\nabla h_\alpha^\theta(X) - \nabla h_\alpha(X)\|^2], \quad X \sim \pi$$

Solution:

$$\theta^* = M^{-1} b$$

$$M = \mathsf{E}_\pi\Big[(\nabla\psi(X))^{\intercal}\nabla\psi(X)\Big]$$
$$b = \mathsf{E}_\pi\Big[(\nabla\psi(X))^{\intercal}\nabla h_\alpha(X)\Big]$$
$$\nabla h_\alpha = \Omega_\alpha \nabla c$$

Adjoint gives causal representation:

$$b_i = \langle \partial^i \psi, \Omega_\alpha \nabla c\rangle$$
$$= \langle \Omega_\alpha^\dagger \partial^i \psi, \nabla c\rangle \quad = \mathsf{E}_\pi[\varphi(t)^{\intercal}\nabla c(X(t))],$$

$$\varphi(t) = \sum_{k=0}^{\infty} \alpha^k [\mathcal{A}(1+t-k)\mathcal{A}(2+t-k)\cdots\mathcal{A}(t)]^{\intercal}\nabla\psi(X(t-k)), \quad t \in \mathbb{Z}$$

Differential Least Squares Temporal Difference Algorithm

∇-LSTD algorithm

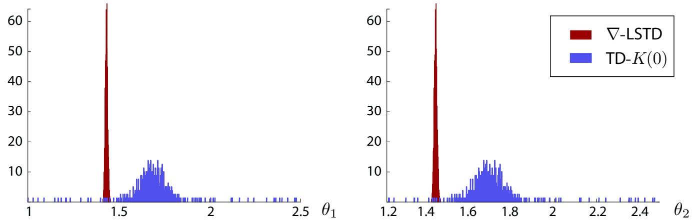
$$
\begin{aligned}
\varphi(t) &= \alpha \mathcal{A}^{\intercal}(t)\varphi(t-1) + \nabla\psi(X(t)) \\
b(t) &= (1-\gamma_t)b(t-1) + \gamma_t \varphi(t)^{\intercal}\nabla c(X(t)) \\
M(t) &= (1-\gamma_t)M(t-1) + \gamma_t \nabla\psi(X(t))\nabla\psi(X(t))^{\intercal} \\
\theta(t) &= M^{-1}(t)b(t)
\end{aligned}
$$

Differential Least Squares Temporal Difference Algorithm

∇-LSTD algorithm

$$\varphi(t) = \alpha \mathcal{A}^T(t)\varphi(t-1) + \nabla\psi(X(t))$$
$$b(t) = (1-\gamma_t)b(t-1) + \gamma_t \varphi(t)^T \nabla c(X(t))$$
$$M(t) = (1-\gamma_t)M(t-1) + \gamma_t \nabla\psi(X(t))\nabla\psi(X(t))^T$$
$$\theta(t) = M^{-1}(t)b(t)$$

Algorithm is amazing:



Conclusions

New Banach space for Markov processes is just right for our goals:

$$\mathcal{D}h = \mathcal{G}(h, \nabla h)$$

Conclusions

New Banach space for Markov processes is just right for our goals:

$$\mathcal{D}h = \mathcal{G}(h, \nabla h)$$

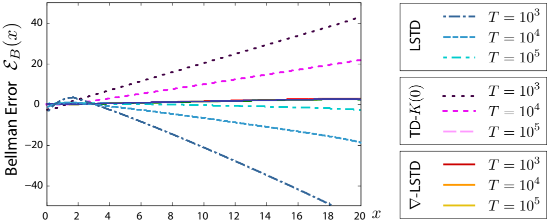
Gaps:
Relationship with Lyapunov exponents remains a mystery.
Needed for a firmer theory for the ∇-LSTD algorithm.

Conclusions

New Banach space for Markov processes is just right for our goals:

$$\mathcal{D}h = \mathcal{G}(h, \nabla h)$$

Gaps:
Relationship with Lyapunov exponents remains a mystery.
Needed for a firmer theory for the ∇-LSTD algorithm.



Thank you!

Selected References

- A. Devraj, I. Kontoyiannis, and S. P. Meyn, "Exponential ergodicity and Lyapunov exponents Part I: Markov chains in discrete time," In preparation, 2016.
- I. Kontoyiannis and S. P. Meyn. "Computable exponential bounds for screened estimation and simulation," *Ann. Appl. Probab.*, 18(4):1491–1518, 2008.
- I. Kontoyiannis and S. P. Meyn. "Approximating a diffusion by a finite-state hidden Markov model," *Stochastic Proc. and their Appl.*, 2016.
- S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*, 2nd ed. Cambridge Mathematical Library, 2009.
- A. M. Devraj and S. P. Meyn. Differential TD learning for value function approximation. In *55th Conference on Decision and Control*, pages 6347–6354, Dec 2016.
- S. P. Meyn, *Control Techniques for Complex Networks*. Cambridge University Press, 2007, pre-publication edition available online.