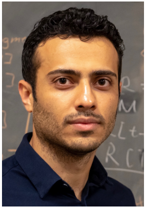
Convergence and sample complexity of gradient methods for the model-free LQR problem

Mihailo Jovanović

viterbi-web.usc.edu/∼mihailo/



Hesam Mohammadi   Armin Zare   Mahdi Soltanolkotabi

NREL Workshop on Resilient Autonomous Energy Systems

Context

- CAN WE LEARN CONTROLLERS?
  - ⋆ complex systems; uncertain environment
  - ⋆ absence of first-principle models
  - ⋆ abundance of data

A case study: LQR problem

minimize	quadratic cost
subject to	linear dynamics

A case study: LQR problem

$$\begin{array}{rl} \text{minimize} & \text{quadratic cost} \\ \text{subject to} & \text{linear dynamics} \end{array}$$

LTI dynamics: $$\frac{\mathrm{d}x(t)}{\mathrm{d}t} \;=\; A\,x(t) \,+\, B\,u(t), \;\; x(0) \;=\; x_0$$

quadratic cost: $$\mathbb{E}_{x_0}\left[\int_0^\infty \big(x^T(t)\,Q\,x(t) \,+\, u^T(t)\,R\,u(t)\big)\,\mathrm{d}t\right]$$

A, B – *model parameters*

Q, $R \succ 0$ – *state and control weights*

- OPTIMAL SOLUTION
  - linear feedback policy

$$u(t) \;=\; -K^{\star} x(t)$$

Riccati-based characterization

$$\begin{aligned} A^T P \;+\; P A \;-\; P B R^{-1} B^T P \;+\; Q \;&=\; 0 \\ K^{\star} \;&=\; R^{-1} B^T P \end{aligned}$$

Model-free reinforcement learning

- LEARN FEEDBACK POLICY BY OPTIMIZING OVER K

$$\underset{K}{\text{minimize}} \quad f(K)$$

- QUESTIONS
  - ⋆ convergence properties (lack of convexity)
  - ⋆ statistical properties (lack of models)

Our contributions

- GRADIENT FLOW DYNAMICS FOR LQR
  - ⋆ **exponential stability:** over the set of stabilizing K

$$\frac{\mathrm{d}K(t)}{\mathrm{d}t} \;=\; -\nabla f(K(t))$$

Our contributions

- Gradient flow dynamics for LQR
  - ⋆ **exponential stability:** over the set of stabilizing K

$$\frac{\mathrm{d}K(t)}{\mathrm{d}t} \;=\; -\nabla f(K(t))$$

Gradient descent for LQR

- ⋆ **linear convergence:** over the set of stabilizing K

$$K^{k+1} \;=\; K^k \;-\; \alpha\,\nabla f(K^k)$$

- RANDOM SEARCH METHOD FOR LQR

convergence and sample complexity

linear convergence:	over the set of stabilizing K
logarithmic complexity:	wrt the reciprocal of accuracy

with high probability

Mohammadi, Zare, Soltanolkotabi, Jovanović, arXiv:1912.11899

IEEE TAC (in press)

Linear convergence of gradient descent

SMOOTHNESS

$$K \in \mathcal{S}_a \;\Rightarrow\; \|\nabla^2 f(K)\|_2 \,\leq\, L$$

GRADIENT DOMINANCE

$$K \in \mathcal{S}_a \;\Rightarrow\; \|\nabla f(K)\|_F^2 \,\geq\, \mu\,(f(K) \,-\, f(K^\star))$$

- SUB-LEVEL SET OF THE LQR COST

$$\mathcal{S}_a \;:=\; \{\text{stabilizing } K \quad f(K) \,\leq\, a$$

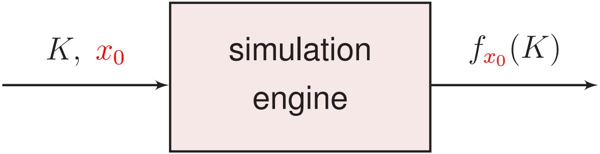
Fazel, Ge, Kakade, Mesbahi, ICML '18

Mohammadi, Zare, Soltanolkotabi, Jovanović, IEEE CDC '19

RANDOM SEARCH METHOD FOR LQR:
convergence and sample complexity

Model-free setup

- GRADIENT IS NOT AVAILABLE
  - ⋆ can only access random function values



  - ⋆ objective function

$$f(K) \;=\; \mathbb{E}_{x_0}[f_{x_0}(K)]$$

x_0 – *sub-Gaussian*

Random search for model-free LQR

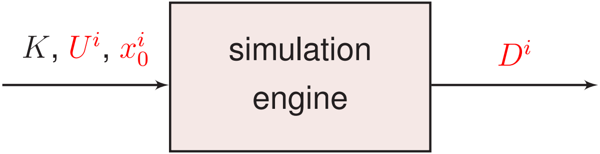
- EMULATE GRADIENT DESCENT

$$K^+ \; = \; K \; - \; \alpha D$$

  - ⋆ gradient estimation using random function values

$$D \; = \; \begin{cases} \dfrac{f_{x_0}(K + rU) \; - \; f_{x_0}(K - rU)}{2r} \, U & \text{two-point} \\[2ex] \dfrac{f_{x_0}(K + rU) \; - \; f_{\hat{x}_0}(K - rU)}{2r} \, U & \text{one-point} \end{cases}$$

U – *random matrices*



⋆ mini-batch gradient estimate

$$D \;=\; \frac{1}{N} \sum_{i\,=\,1}^{N} D^i$$

N – *number of samples*

Main result: linear rate; log complexity

- THEOREM

  Random search LQR achieves a desired accuracy ϵ, i.e.,

$$f(K^T) \; - \; f(K^\star) \; \leq \; \epsilon$$

with high probability, if

number of iterations	$T \; = \; O\left(\log\left(1/\epsilon\right)\right)$
number of samples (per iteration)	$N \; = \; c\,n\,(\log n)^6$

n – number of states

two-point estimate

Mohammadi, Zare, Soltanolkotabi, Jovanović, arXiv:1912.11899

State-of-the-art for random search LQR

polynomial sample complexity

sub-linear convergence rate

decreasing stepsize

for ϵ-accuracy	*one-point*	*two-point*
iterations T	$O(\epsilon^{-2})$	$O(\epsilon^{-1})$
samples N	1	1
stepsize α	$O(\epsilon^{2})$	$O(\epsilon)$

Malik, Pananjady, Bhatia, Khamaru, Bartlett, Wainwright, JMLR '20

- ONE-POINT MINI-BATCH ESTIMATE
  - polynomial complexity, linear rate, fixed stepsize

Fazel, Ge, Kakade, Mesbahi, ICML '18

- Our approach
  - DON'T control gradient estimation error

key in our proof

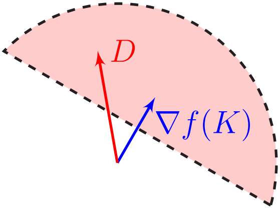
gradient estimate concentrates with high probability
(when projected to the direction of the gradient)

two-point estimate

Approximate gradient descent

$$\langle D, \nabla f(K) \rangle \;\geq\; \theta_1 \, \|\nabla f(K)\|_F^2$$

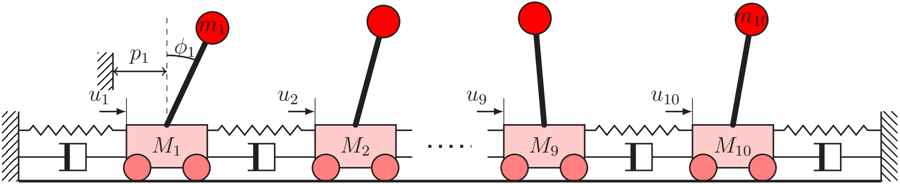
$$\|D\|_F^2 \;\leq\; \theta_2 \, \|\nabla f(K)\|_F^2$$



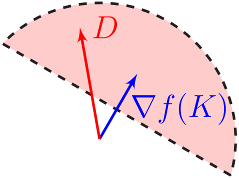
LINEAR CONVERGENCE WITH FIXED STEPSIZE

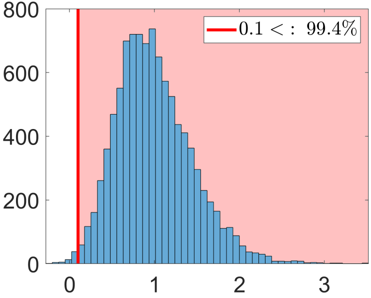
$$f(K^k) \,-\, f(K^\star) \;\leq\; \left(1 \,-\, \frac{\theta_1^2 \, \mu}{\theta_2 \, L}\right)^k \big(f(K^0) \,-\, f(K^\star)\big)$$

An example

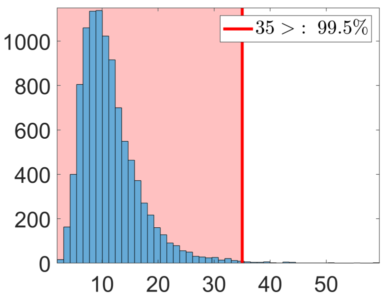


Occurrence with high probability



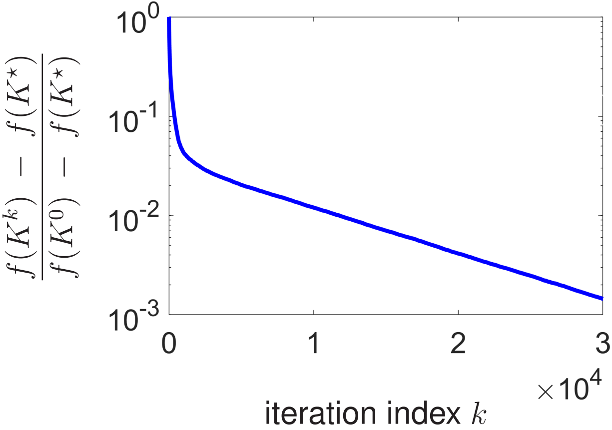


$$\frac{\langle D, \nabla f(K)\rangle}{\|\nabla f(K)\|_F^2} \geq \underbrace{0.1}_{\theta_1}$$



$$\frac{\|D\|_F^2}{\|\nabla f(K)\|_F^2} \leq \underbrace{35}_{\theta_2}$$

Linear convergence



GRADIENT FLOW DYNAMICS FOR LQR:
global exponential stability

Challenge: lack of convexity

$$f(K) \;=\; \begin{cases} \operatorname{trace}\big((Q \,+\, K^T R\, K)\, X(K)\big)\,, & K \,\in\, \mathcal{S} \\ \infty, & \text{otherwise} \end{cases}$$

$$X(K) \;=\; \int_0^\infty \mathbb{E}_{x_0}\big(x(t)x^T(t)\big)\,\mathrm{d}t \;=\; \int_0^\infty \mathrm{e}^{(A\,-\,B\,K)\,t}\,\Omega\,\mathrm{e}^{(A\,-\,B\,K)^T t}\,\mathrm{d}t$$

$$\Omega \;:=\; \mathbb{E}\left(x_0 x_0^T\right) \;\succ\; 0$$

Challenge: lack of convexity

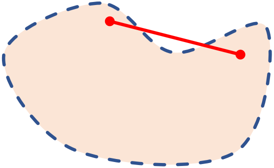
$$f(K) \;=\; \begin{cases} \operatorname{trace}\left((Q + K^T R\, K)\, X(K)\right), & K \in \mathcal{S} \\ \infty, & \text{otherwise} \end{cases}$$

$$X(K) \;=\; \int_0^\infty \mathbb{E}_{x_0}\left(x(t)x^T(t)\right) \mathrm{d}t \;=\; \int_0^\infty \mathrm{e}^{(A - B\,K)\,t}\,\Omega\,\mathrm{e}^{(A - B\,K)^T t}\,\mathrm{d}t$$

$$\Omega \;:=\; \mathbb{E}\left(x_0 x_0^T\right) \;\succ\; 0$$

- SET OF STABILIZING FEEDBACK GAINS

$\mathcal{S} \;:=\; \{K \;\; A \,-\, B\,K$ is Hurwitz



Objective

ANALYSIS OF GRADIENT FLOW/DESCENT

$$\frac{\mathrm{d}K(t)}{\mathrm{d}t} = -\nabla f(K(t))$$

$$K^{+} = K - \alpha \nabla f(K)$$

Objective

ANALYSIS OF GRADIENT FLOW/DESCENT

$$\frac{\mathrm{d}K(t)}{\mathrm{d}t} \;=\; -\nabla f(K(t))$$

$$K^{+} \;=\; K \;-\; \alpha \nabla f(K)$$

- GLOBAL CONVERGENCE PROPERTIES
  - ⋆ discrete-time LQR

    Fazel, Ge, Kakade, Mesbahi, ICML '18
  - ⋆ continuous-time LQR

    Mohammadi, Zare, Soltanolkotabi, Jovanović, IEEE CDC '19

Key known properties

unique critical point: $\nabla f(K^\star) \;=\; 0$

coercivity: $\lim_{K \,\to\, \partial\mathcal{S}/\infty} f(K) \;=\; +\infty$

Toivonen, Int. J. Control '85

Gradient flow dynamics

$$\frac{\mathrm{d}K}{\mathrm{d}t} \;=\; -\,\nabla f(K)$$

- OBJECTIVE ERROR IS A LYAPUNOV FUNCTION

$$K \,\in\, \mathcal{S}\backslash\{K^\star\} \quad \& \quad \text{coercivity of } f$$

$$\Downarrow$$

$$V(K) \;=\; f(K) \,-\, f(K^\star) \;>\; 0$$

$$\dot{V}(K) \;=\; -\,\|\nabla f(K)\|_F^2 \quad <\; 0$$

Gradient flow dynamics

$$\frac{\mathrm{d}K}{\mathrm{d}t} \;=\; -\,\nabla f(K)$$

- Objective error is a Lyapunov function

$$K \;\in\; \mathcal{S}\backslash\{K^\star\} \quad \& \quad \text{coercivity of } f$$

$$\Downarrow$$

$$\begin{aligned} V(K) \;&=\; f(K) \;-\; f(K^\star) \;>\; 0 \\ \dot{V}(K) \;&=\; -\,\|\nabla f(K)\|_F^2 \quad\;\; <\; 0 \end{aligned}$$

$$\Downarrow$$

$$\lim_{t\,\to\,\infty} K(t) \;=\; K^\star \;\; \text{for all } K \in \mathcal{S}$$

asymptotic stability on $\mathcal{S}$

Exponential stability

- POLYAK-LOJASIEWICZ CONDITION
  - ⋆ gradient dominance

$$\|\nabla f(K)\|_F^2 \;\geq\; \rho\,(f(K) \,-\, f(K^\star))$$

$$\Downarrow$$

$$\dot{V}(K) \;\leq\; -\rho\,V(K)$$

Exponential stability

- Polyak-Lojasiewicz condition

  - ⋆ gradient dominance

$$\|\nabla f(K)\|_F^2 \;\geq\; \rho\,(f(K) \,-\, f(K^\star))$$

$$\Downarrow$$

$$\dot{V}(K) \;\leq\; -\rho\, V(K)$$

  - ⋆ an example: strongly convex objective function

$$\nabla^2 f(K) \;\succ\; \frac{\rho}{2}\, I$$

Convex re-parameterization

$$\begin{array}{ll} \underset{X,\,K}{\text{minimize}} & \text{trace}\left((Q + K^T R\, K) X\right) \\ \text{subject to} & (A - BK)X + X(A - BK)^T + \Omega = 0 \\ & X \succ 0 \end{array}$$

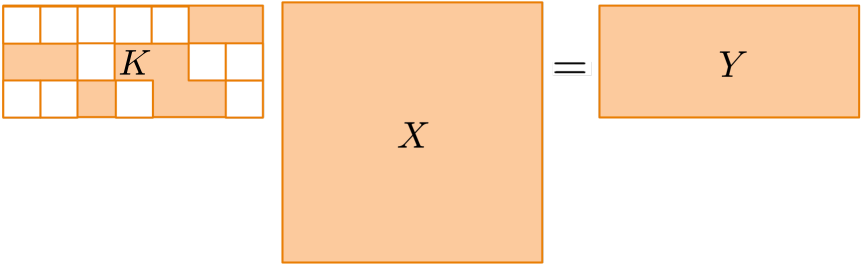
closed-loop stability

Convex re-parameterization

$$\begin{array}{ll} \underset{X,\,K}{\text{minimize}} & \text{trace}\left((Q + K^T R K)X\right) \\ \text{subject to} & (A - BK)X + X(A - BK)^T + \Omega = 0 \\ & X \succ 0 \end{array}$$

closed-loop stability

⋆ **change of variables:** $KX =: Y$



$$\underset{X,Y}{\text{minimize}} \quad \overbrace{\text{trace}\,(Q\,X) \; + \; \text{trace}\left(R\,Y\,X^{-1}\,Y^T\right)}^{h(X,Y)}$$

$$\text{subject to} \quad (A\,X \, - \, B\,Y) + (A\,X \, - \, B\,Y)^T + \Omega \, = \, 0$$

$$X \, \succ \, 0$$

Schur complement $\Rightarrow$ SDP characterization

Feron, Balakrishnan, Boyd, El Ghaoui '92

Our approach

⋆ express X in terms of Y

$$X(Y) \;=\; \mathcal{A}^{-1}(\mathcal{B}(Y) \;-\; \Omega) \;\succ\; 0$$ affine function of Y

$$\mathcal{A}(X) \;=\; A\,X \;+\; X A^T$$

$$\mathcal{B}(Y) \;=\; B\,Y \;+\; Y^T B^T$$

write $f(K)$ as a convex function of Y

$$h(Y) \;=\; \operatorname{trace}\,(Q\,X(Y)) \;+\; \operatorname{trace}\left(R\,Y X^{-1}(Y)\,Y^T\right)$$

GRADIENT FLOW OVER THE CONVEX LANDSCAPE

$$\frac{\mathrm{d}Y(t)}{\mathrm{d}t} = -\nabla h(Y(t))$$

Key property

- STRONG CONVEXITY OVER THE SUB-LEVEL SETS

  ⋆ Hessian

$$\langle G, \nabla^2 h(Y;G) \rangle \;=\; 2 \quad R^{\frac{1}{2}}(G \;-\; K\mathcal{A}^{-1}(\mathcal{B}(G)))X^{-\frac{1}{2}} \quad {}^{2}_{F}$$

$$\Downarrow$$

$$G, \nabla^2 h(Y;G) \quad \geq \; \rho(a)\|G\|_F^2 \;\; \text{for all } Y \in \tilde{\mathcal{S}}_a$$

  ⋆ sub-level sets of h

$$\tilde{\mathcal{S}}_a \;=\; \{Y \in \mathbb{R}^{m\times n} \mid X(Y) \succ 0, \; h(Y) \leq a\}$$

EXPONENTIAL STABILITY OF $\dot{Y} = -\nabla h(Y)$

for any initial condition $Y(0) \in \tilde{\mathcal{S}}_a$

$$h(Y(t)) - h(Y^\star) \leq (h(Y(0)) - h(Y^\star))\, \mathrm{e}^{-\rho(a)t}$$

EXPONENTIAL STABILITY OF $\dot{Y} = -\nabla h(Y)$

for any initial condition $Y(0) \in \tilde{\mathcal{S}}_a$

$$h(Y(t)) - h(Y^\star) \leq (h(Y(0)) - h(Y^\star))\, \mathrm{e}^{-\rho(a)t}$$

properties of $\dot{K} = -\nabla f(K)$?

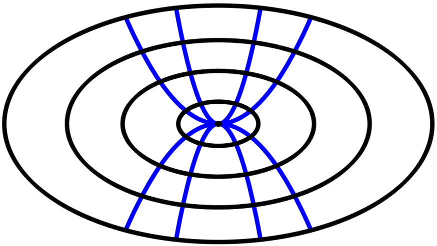
Relating convex and non-convex landscapes

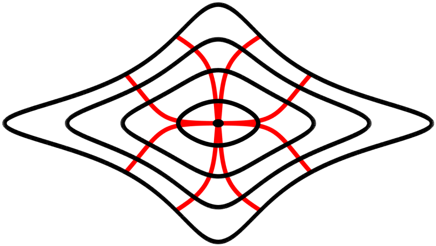
convex function $h(Y)$ $\quad = \quad$ non-convex function $f(K)$

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = -\nabla h(Y) \qquad\qquad \frac{\mathrm{d}K}{\mathrm{d}t} = -\nabla f(K)$$

$$\downarrow \qquad\qquad\qquad\qquad \downarrow$$

$$Y(t) \qquad\qquad\qquad\qquad K(t)$$





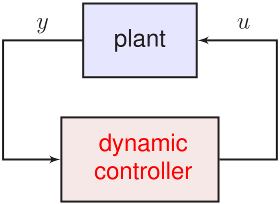
Mohammadi, Zare, Soltanolkotabi, Jovanović, arXiv:1912.11899

Ongoing effort

- SPARSITY-PROMOTING RL

$$
\begin{bmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \\ u_4(t) \\ u_5(t) \end{bmatrix}
= - \underbrace{\begin{bmatrix} * & * & & & \\ * & * & & & * \\ & * & * & * & \\ & & * & * & \\ & * & & * & * \end{bmatrix}}_{K}
\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \\ x_5(t) \end{bmatrix}
$$

- OUTPUT-FEEDBACK RL



References

DISCRETE-TIME LQR

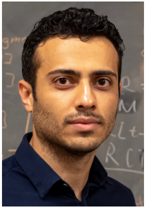
- ⋆ *Fazel, Ge, Kakade, Mesbahi, ICML '18*
- ⋆ *Malik et al., JMLR '20*
- ⋆ *Mohammadi, Soltanolkotabi, Jovanović, L-CSS '21*

CONTINUOUS-TIME LQR

- ⋆ *Mohammadi, Zare, Soltanolkotabi, Jovanović, IEEE CDC '19*
- ⋆ *Mohammadi, Soltanolkotabi, Jovanović, L4DC '20*
- ⋆ *Mohammadi, Zare, Soltanolkotabi, Jovanović*, arXiv:1912.11899

  IEEE TAC (in press)

Acknowledgements


Hesam


Armin


Mahdi

- SUPPORT
  - ⋆ AFOSR FA9550-16-1-0009
  - ⋆ NSF ECCS-1809833