

Learning the Globally Optimal Distributed LQ Regulator

• Luca Furieri¹, Yang Zheng², Maryam Kamgarpour¹

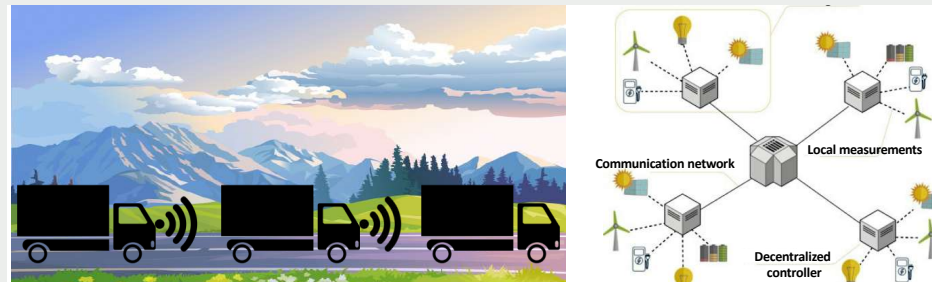
¹Automatic Control Laboratory, ETH Zurich; ²School of Engineering and Applied Sciences, Harvard University

ETH zürich



HARVARD
UNIVERSITY

1 Introduction



Goal: Operate safety-critical large-scale dynamical systems, such as the continent-wide power grid and traffic networks of autonomous vehicles.

Challenges: 1) Models are not available, and
2) Multiple myopic decision-makers.

A theory of distributed, learning-based control is needed.

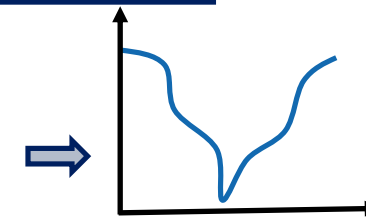
3 Results and Discussion

Properties of the LQ cost for distributed control

- The policy \mathbf{K} is constrained to a *sparsity subspace* \mathcal{K}
- $J(\mathbf{K})$ is generally **non-convex** on \mathcal{K} ! Issues with Gradient Descent (GD) methods. (saddle-points, local mins...)
- Quadratic Invariance (QI):** there exists transformation $\mathbf{K} = h(\mathbf{Q})$ such that $J(h(\mathbf{Q}))$ is strongly convex on $\mathbf{Q} \in \mathcal{K}$, if and only if QI holds, that is:
 $\mathbf{K}\mathbf{G}\mathbf{K} \in \mathcal{K}, \forall \mathbf{K} \in \mathcal{K}. (\text{QI})$

What is the role of QI for learning?

Theorem 1. Assume QI. Then $J(\mathbf{K})$ has a *gradient-dominance* constant for any sublevel set.



Proof sketch: Given QI, $J(h(\cdot))$ is strongly-convex and has gradient-dominance constant. Since $h(\cdot)$ is invertible, sublevel sets of $J(\cdot)$ are bounded. Last, Jacobian of $J(\cdot)$ is bounded on sublevel sets.

Theorem 2. QI is not necessary for gradient dominance.

Proof sketch: Explicit *non-QI* example, where A_t is dense, \mathbf{K} is diagonal and $J(\mathbf{K})$ is convex on diagonal subspace.

Idea: Let us 1) estimate the gradient and 2) descend it.
Gradient-dominance = convergence to global optimum!

Theorem 3. Assume QI. In the **Algorithm** select *stepsize* η and *smoothing radius* r as

$$\eta \leq O\left(\frac{\epsilon \delta^3 r^2}{w^2 J(\mathbf{K}_0)^2}\right), \quad r \leq O(\sqrt{\delta \epsilon}),$$

and the iterations as $T = O(\eta^{-1} \log(\delta^{-1} \epsilon^{-1}))$. Then

$$J(\mathbf{K}_T) - J^* \leq \epsilon,$$

with probability greater than $1 - \delta$.

Proof sketch: martingale argument similar to [2].
Trade-off: larger T = larger uncertainty = closer to optimum.

Zeroth-order RL Algorithm

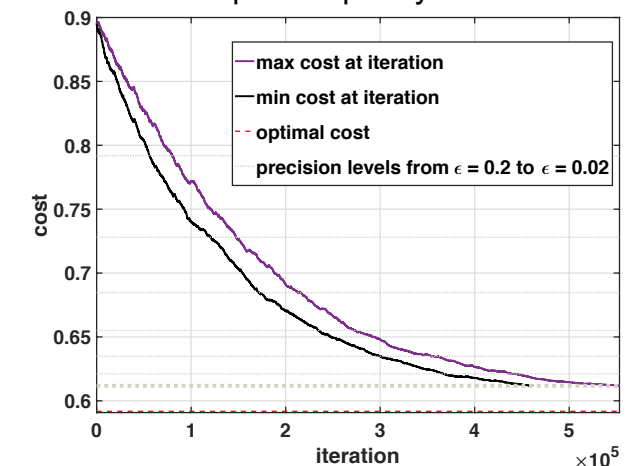
Algorithm 1 Model-free learning of distributed controllers

```

1: Input:  $z_0$ , number of iterations  $T$ , stepsize  $\eta > 0$  and smoothing radius  $r > 0$ .
2: for  $i = 0, \dots, T-1$  do
3:   Sample  $u \sim \text{Unif}(\mathbb{S}_r)$ , let nature “choose” disturbances  $\delta_0 \sim \mathcal{D}_{\delta_0}$ ,  $w_t \sim \mathcal{D}_w$  for all  $t = 0, \dots, N-1$ ,  $v_t \sim \mathcal{D}_v$  for all  $t = 0, \dots, N$ .
4:   Apply the control policy  $\hat{\mathbf{u}} = \text{vec}^{-1}[P(z_i + u)]\hat{\mathbf{y}}$  and store the resulting trajectories  $\hat{\mathbf{y}}, \hat{\mathbf{u}}$ .
5:   Compute  $\hat{f} = \hat{\mathbf{y}}^\top \text{blkdg}(M_0, \dots, M_N)\hat{\mathbf{y}} + \hat{\mathbf{u}}^\top \text{blkdg}(R_0, \dots, R_{N-1})\hat{\mathbf{u}}$  and  $\hat{\nabla} f = \hat{f}_{\frac{d}{r}} u$ .
6:    $z_{i+1} \leftarrow z_i - \eta \hat{\nabla} f$ .
7: end for
8: return  $\mathbf{K}_T = \text{vec}^{-1}(P z_T)$ .
    
```

4 Experiments

We validate the sample-complexity results.



5 Conclusions

QI (and beyond!) = RL of optimal distributed controllers!

Extensions: infinite-horizon, distributed learning, safety...

6 References

- “Learning the Globally Optimal Distributed LQ Regulator”, L. Furieri, Y. Zheng, M. Kamgarpour, 2020
- “Derivative-Free Methods for Policy Optimization: Guarantees for Linear Quadratic Systems”, D. Malik, A. Pananjady, K. Bhatia, K. Khamaru, P. Bartlett, M. Wainwright, 2019
- “Global Convergence of Policy Gradient Methods for the Linear Quadratic Regulator”, M. Fazel, R. Ge, S. Kaked, M. Mesbahi, 2018
- “First Order Methods For Globally Optimal Distributed Controllers Beyond Quadratic Invariance”, L. Furieri, M. Kamgarpour, ACC 2020

2 Set-Up

Consider a linear system

$$x_{t+1} = A_t x_t + B_t u_t + w_t, \quad y_t = C_t x_t + v_t$$

in finite-horizon N , where $x_0, w_t, v_t \sim D$ are bounded, and the **dynamics** A_t, B_t, C_t are **unknown**. Consider policies

$$u_t = K_{t0} y_0 + K_{t1} y_1 + \dots + K_{tt} y_t$$

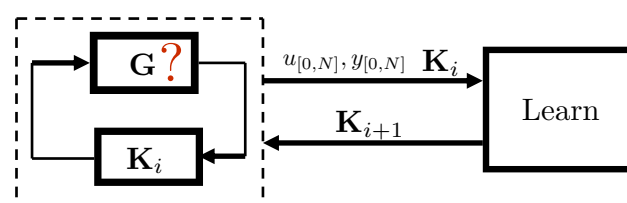
for every $t = 0, \dots, N$, where K_{ij} are the decision variables.

• **Distributed Control:** some outputs not available to u_t !

The K'_{ij} s must have a *sparsity pattern*: e.g. $K'_{ij} = \begin{bmatrix} * & 0 & * \\ 0 & * & * \end{bmatrix}$

• **LQ Cost:** $J(\mathbf{K}) := \mathbb{E}_{w,v} \left[\sum_{t=0}^{N-1} (y_t^\top M_t y_t + u_t^\top R_t u_t) + y_N^\top M_N y_N \right]$

where \mathbf{K} stacks the K'_{ij} s together.



Converge to globally **optimal distributed** controller?