

Min Max Generalization for Deterministic Batch Mode Reinforcement Learning

Raphael Fonteneau^{1,2}

The work presented in this talk was done in collaboration with:

Bernard Boiglot¹,
Damien Ernst¹,
Quentin Louveaux¹
Susan A. Murphy³
Louis Wehenkel¹

¹ Department of EECS, University of Liège, Belgium

² Inria Lille – Nord Europe, France

³ Department of Statistics, University of Michigan, Ann Arbor, USA

Sequel – Inria Lille – Nord Europe, France – 15 juin 2012

Goal



Author: ArcCan – Wikipedia Commons

How to safely control a deterministic system living in a continuous state space given the knowledge of:

- A batch collection of trajectories of the system,
- The maximal variations of the system (upper bounds on the Lipschitz constants).

Menu

Introduction

I Direct approach

- CGRL algorithm

II Reformulation of the original problem

- 2 relaxations schemes in the two-stage case

III Comparison of the 3 proposed solutions

Conclusions and future work

Formalization

- Deterministic dynamics: $x_{t+1} = f(x_t, u_t) \quad t = 0, \dots, T-1,$
- Deterministic reward function: $r_t = \rho(x_t, u_t) \in \mathbb{R}$
- Fixed initial state: $x_0 \in \mathcal{X}$
- Continuous state space, finite action space: $\mathcal{X} \subset \mathbb{R}^d \quad \mathcal{U} = \{u^{(1)}, \dots, u^{(m)}\}$
- Return of a sequence of actions:

$$\forall (u_0, \dots, u_{T-1}) \in \mathcal{U}^T, \quad J_T^{(u_0, \dots, u_{T-1})} = \sum_{t=0}^{T-1} \rho(x_t, u_t)$$

- Optimal return:

$$J_T^* = \max_{(u_0, \dots, u_{T-1}) \in \mathcal{U}^T} J_T^{(u_0, \dots, u_{T-1})}$$

The "batch" setting

- Dynamics and reward function are **unknown**
- For all actions $u \in \mathcal{U}$, a set of one-step transitions is given:

$$\mathcal{F}^{(u)} = \left\{ \left(x^{(u),k}, r^{(u),k}, y^{(u),k} \right) \right\}_{k=1}^{n^{(u)}}$$

$$y^{(u),k} = f \left(x^{(u),k}, u \right) \quad r^{(u),k} = \rho \left(x^{(u),k}, u \right)$$

$$\forall u \in \mathcal{U}, \quad n^{(u)} > 0$$

- We note:

$$\mathcal{F} = \mathcal{F}^{(1)} \cup \dots \cup \mathcal{F}^{(m)}$$

Lipschitz continuity

- The system dynamics and the reward function are Lipschitz continuous :

$$\begin{aligned} \forall (x, x') \in \mathcal{X}^2, \forall u \in \mathcal{U}, \quad & \|f(x, u) - f(x', u)\| \leq L_f \|x - x'\| \\ & |\rho(x, u) - \rho(x', u)| \leq L_\rho \|x - x'\| \end{aligned}$$

where $\|\cdot\|$ is the Euclidean norm over the state space.

- We assume that two constants L_f and L_ρ satisfying the above inequalities are known.

Compatible environments

- Compatible dynamics and reward functions:

$$\mathcal{L}_{\mathcal{F}}^f = \left\{ f' : \mathcal{X} \times \mathcal{U} \rightarrow \mathcal{X} \mid \left\{ \begin{array}{l} \forall x', x'' \in \mathcal{X}, \forall u \in \mathcal{U} , \\ \|f'(x', u) - f'(x'', u)\| \leq L_f \|x' - x''\| , \\ \forall k \in \{1, \dots, n^{(u)}\}, f'(x^{(u),k}, u) = f(x^{(u),k}, u) = y^{(u),k} \end{array} \right. \right\}$$

$$\mathcal{L}_{\mathcal{F}}^{\rho} = \left\{ \rho' : \mathcal{X} \times \mathcal{U} \rightarrow \mathbb{R} \mid \left\{ \begin{array}{l} \forall x', x'' \in \mathcal{X}, \forall u \in \mathcal{U} , \\ |\rho'(x', u) - \rho'(x'', u)| \leq L_{\rho} \|x' - x''\| , \\ \forall k \in \{1, \dots, n^{(u)}\}, \rho'(x^{(u),k}, u) = \rho(x^{(u),k}, u) = r^{(u),k} \end{array} \right. \right\}$$

Compatible environments

- Compatible dynamics and reward functions:

$$\mathcal{L}_{\mathcal{F}}^f = \left\{ f' : \mathcal{X} \times \mathcal{U} \rightarrow \mathcal{X} \mid \begin{cases} \forall x', x'' \in \mathcal{X}, \forall u \in \mathcal{U}, \\ \|f'(x', u) - f'(x'', u)\| \leq L_f \|x' - x''\|, \\ \forall k \in \{1, \dots, n^{(u)}\}, f'(x^{(u),k}, u) = f(x^{(u),k}, u) = y^{(u),k} \end{cases} \right\}$$

$$\mathcal{L}_{\mathcal{F}}^{\rho} = \left\{ \rho' : \mathcal{X} \times \mathcal{U} \rightarrow \mathbb{R} \mid \begin{cases} \forall x', x'' \in \mathcal{X}, \forall u \in \mathcal{U}, \\ |\rho'(x', u) - \rho'(x'', u)| \leq L_{\rho} \|x' - x''\|, \\ \forall k \in \{1, \dots, n^{(u)}\}, \rho'(x^{(u),k}, u) = \rho(x^{(u),k}, u) = r^{(u),k} \end{cases} \right\}$$

- Return of a sequence of actions under a compatible environment:

$$J_{T, (f', \rho')}^{(u_0, \dots, u_{T-1})} = \sum_{t=0}^{T-1} \rho'(x'_t, u_t) \quad \begin{aligned} x'_{t+1} &= f'(x'_t, u_t), \quad \forall t \in \{0, \dots, T-1\} \\ x'_0 &= x_0 \end{aligned}$$

Min max approach to generalization

- Worst possible return of a given sequence of actions:

$$L_T^{(u_0, \dots, u_{T-1})}(\mathcal{F}) = \min_{(f', \rho') \in \mathcal{L}_{\mathcal{F}}^f \times \mathcal{L}_{\mathcal{F}}^p} J_{T, (f', \rho')}^{(u_0, \dots, u_{T-1})}$$

Min max approach to generalization

- Worst possible return of a given sequence of actions:

$$L_T^{(u_0, \dots, u_{T-1})}(\mathcal{F}) = \min_{(f', \rho') \in \mathcal{L}_{\mathcal{F}}^f \times \mathcal{L}_{\mathcal{F}}^p} J_{T, (f', \rho')}^{(u_0, \dots, u_{T-1})}$$

Our objective:

$$(u_0^*, \dots, u_{T-1}^*) \in \arg \max_{(u_0, \dots, u_{T-1}) \in \mathcal{U}^T} L_T^{(u_0, \dots, u_{T-1})}(\mathcal{F})$$

Menu

Introduction

I Direct approach

- CGRL algorithm

II Reformulation of the original problem

- 2 relaxations schemes in the two-stage case

III Comparison of the 3 proposed solutions

Conclusions and future work

Bounds from a sequence of transitions

- For a given sequence of actions, one can compute from a sequence of one-step transitions which is compatible with the sequence of actions a lower-bound on the "worst possible return" of the sequence of actions:

Lemma *Let $(u_0, \dots, u_{T-1}) \in \mathcal{U}^T$ be a sequence of actions.*

Let $\tau = [(x^{(u_t), k_t}, r^{(u_t), k_t}, y^{(u_t), k_t})]_{t=0}^{T-1} \in \mathcal{F}_{(u_0, \dots, u_{T-1})}^T$.

Then,

$$B(\mathcal{F}, \tau) \leq L_T^{(u_0, \dots, u_{T-1})}(\mathcal{F}) ,$$

with

$$B(\mathcal{F}, \tau) \doteq \sum_{t=0}^{T-1} \left[r^{(u_t), k_t} - L_{Q_{T-t}} \left\| y^{(u_{t-1}), k_{t-1}} - x^{(u_t), k_t} \right\| \right] ,$$

$$y^{(u_{-1}), k_{-1}} = x_0 ,$$

$$L_{Q_{T-t}} = L_\rho \sum_{i=0}^{T-t-1} (L_f)^i .$$

Maximal lower bound

- One can define a best lower bound over all possible sequences of transitions:

$$B_{CGRL}^{(u_0, \dots, u_{T-1})}(\mathcal{F}) = \max_{\tau \in \mathcal{F}_{(u_0, \dots, u_{T-1})}^T} B(\mathcal{F}, \tau)$$

Maximal lower bound

- One can define a best lower bound over all possible sequences of transitions:

$$B_{CGRL}^{(u_0, \dots, u_{T-1})}(\mathcal{F}) = \max_{\tau \in \mathcal{F}_{(u_0, \dots, u_{T-1})}^T} B(\mathcal{F}, \tau)$$

- Such a bound is "tight" w.r.t. the dispersion of the batch collection of data:

Theorem

$$\exists C > 0 : \forall (u_0, \dots, u_{T-1}) \in \mathcal{U}^T, J_T^{(u_0, \dots, u_{T-1})} - B_{CGRL}^{(u_0, \dots, u_{T-1})}(\mathcal{F}) \leq C\alpha^*(\mathcal{F})$$

where $\alpha^*(\mathcal{F})$ is the smallest α such that

$$\forall u \in \mathcal{U}, \sup_{x \in \mathcal{X}} \min_{k \in \{1, \dots, n^{(u)}\}} \|x^{(u),k} - x\| \leq \alpha$$

The CGRL algorithm

- One can define a best lower bound over all possible sequences of transitions:

$$(u_0^*, \dots, u_{T-1}^*) \in \arg \max_{(u_0, \dots, u_{T-1}) \in \mathcal{U}^T} B_{CGRL}^{(u_0, \dots, u_{T-1})}(\mathcal{F})$$

The CGRL algorithm

- One can define a best lower bound over all possible sequences of transitions:

$$(u_0^*, \dots, u_{T-1}^*) \in \arg \max_{(u_0, \dots, u_{T-1}) \in \mathcal{U}^T} B_{CGRL}^{(u_0, \dots, u_{T-1})}(\mathcal{F})$$

- Finding such a sequence can be reformulated as a shortest path problem, and the sequence can be found without enumerating all sequences. This is what the CGRL algorithm does.

The CGRL algorithm

- One can define a best lower bound over all possible sequences of transitions:

$$(u_0^*, \dots, u_{T-1}^*) \in \arg \max_{(u_0, \dots, u_{T-1}) \in \mathcal{U}^T} B_{CGRL}^{(u_0, \dots, u_{T-1})}(\mathcal{F})$$

- Finding such a sequence can be reformulated as a shortest path problem, and the sequence can be found without enumerating all sequences. This is what the CGRL algorithm does.

Properties of the CGRL algorithm

- The CGRL solution converges towards an optimal sequence when the dispersion of the sample of transitions converges towards zero.

The CGRL algorithm

- One can define a best lower bound over all possible sequences of transitions:

$$(u_0^*, \dots, u_{T-1}^*) \in \arg \max_{(u_0, \dots, u_{T-1}) \in \mathcal{U}^T} B_{CGRL}^{(u_0, \dots, u_{T-1})}(\mathcal{F})$$

- Finding such a sequence can be reformulated as a shortest path problem, and the sequence can be found without enumerating all sequences. This is what the CGRL algorithm does.

Properties of the CGRL algorithm

- The CGRL solution converges towards an optimal sequence when the dispersion of the sample of transitions converges towards zero,
- If an "optimal trajectory" is available in the batch sample, that CGRL **will** return such an optimal sequence,

The CGRL algorithm

- One can define a best lower bound over all possible sequences of transitions:

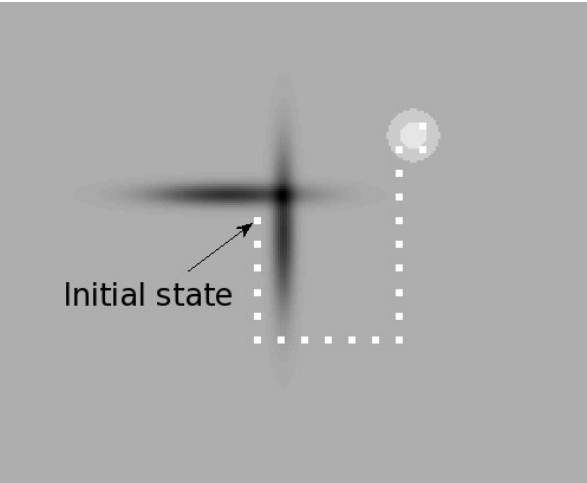
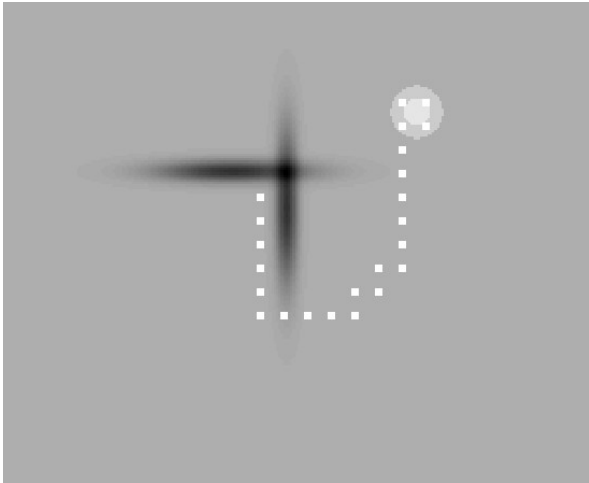
$$(u_0^*, \dots, u_{T-1}^*) \in \arg \max_{(u_0, \dots, u_{T-1}) \in \mathcal{U}^T} B_{CGRL}^{(u_0, \dots, u_{T-1})}(\mathcal{F})$$

- Finding such a sequence can be reformulated as a shortest path problem, and the sequence can be found without enumerating all sequences. This is what the CGRL algorithm does.

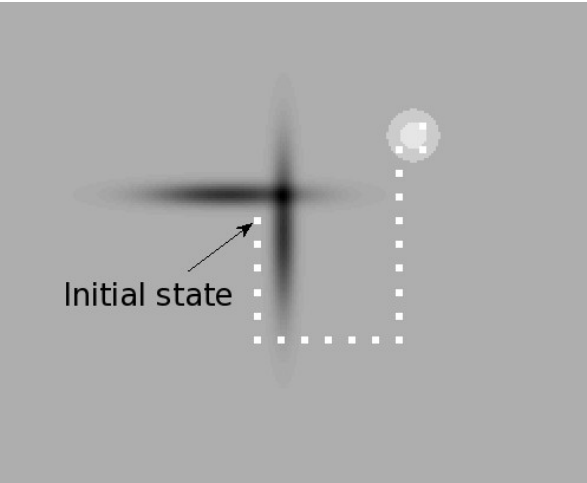
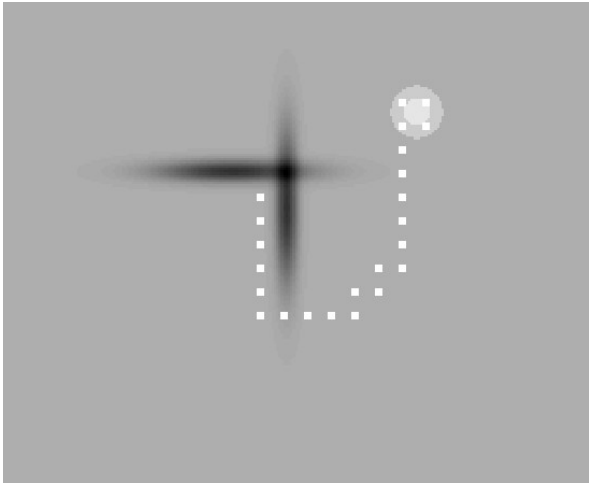
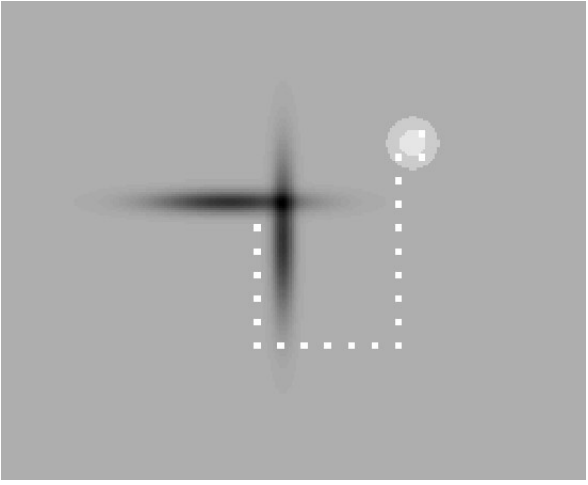
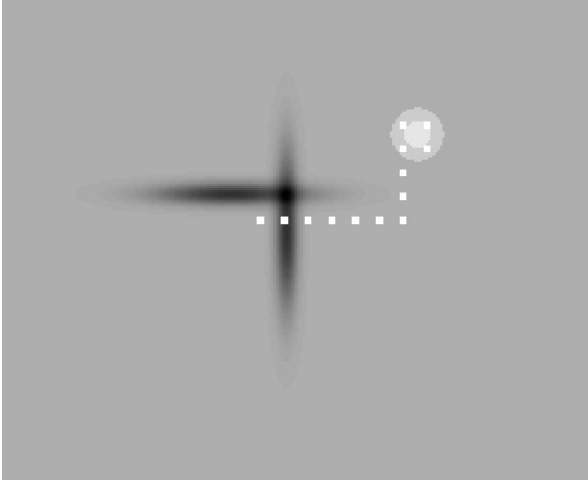
Properties of the CGRL algorithm

- The CGRL solution converges towards an optimal sequence when the dispersion of the sample of transitions converges towards zero.
- If an "optimal trajectory" is available in the batch sample, that CGRL **will** return such an optimal sequence.
- Computational complexity: quadratic w.r.t the size of the batch collection of transitions.

The CGRL algorithm - illustration

The puddle world	CGRL	FQI (Fitted Q Iteration)
<p>The state space is uniformly covered by the sample</p>	 <p>The CGRL visualization shows a grid world with a dark cross-shaped obstacle. A dotted path starts from an 'Initial state' (indicated by an arrow) and moves towards a goal state (a small circle). The path is composed of discrete points, illustrating the state space coverage.</p>	 <p>The FQI visualization shows the same grid world environment. A dotted path starts from the initial state and moves towards the goal state, illustrating the state space coverage for the FQI algorithm.</p>

The CGRL algorithm - illustration

The puddle world	CGRL	FQI (Fitted Q Iteration)
The state space is uniformly covered by the sample	 <p>The CGRL visualization shows a grid world with a central black cross representing a puddle. A dotted path starts from an 'Initial state' (indicated by an arrow) and moves towards a goal state (a small circle). The path is composed of many small squares, indicating a dense and uniform coverage of the state space.</p>	 <p>The FQI visualization shows the same grid world with a central black cross. A dotted path starts from an initial state and moves towards a goal state. The path is composed of fewer squares compared to the CGRL visualization, indicating a sparser coverage of the state space.</p>
Information about the Puddle area is removed	 <p>The CGRL visualization shows the same grid world with a central black cross. A dotted path starts from an initial state and moves towards a goal state. The path is composed of many small squares, indicating a dense and uniform coverage of the state space, even in the area around the puddle.</p>	 <p>The FQI visualization shows the same grid world with a central black cross. A dotted path starts from an initial state and moves towards a goal state. The path is composed of fewer squares compared to the CGRL visualization, indicating a sparser coverage of the state space, particularly in the area around the puddle.</p>

Menu

Introduction

I Direct approach

- CGRL algorithm

II Reformulation of the original problem

- 2 relaxations schemes in the two-stage case

III Comparison of the 3 proposed solutions

Conclusions and future work

Reformulation of min max problem

Theorem (Equivalence). *Computing the worst possible return of the sequence of actions (u_0, \dots, u_{T-1}) is equivalent to solving the problem:*

$(\mathcal{P}_T(\mathcal{F}, L_f, L_\rho, x_0, u_0, \dots, u_{T-1})) :$

$$\begin{array}{ll} \min & \sum_{t=0}^{T-1} \hat{\mathbf{r}}_{\mathbf{t}}, \\ \hat{\mathbf{r}}_{\mathbf{0}} \quad \dots \quad \hat{\mathbf{r}}_{\mathbf{T}-1} \in \mathbb{R} & \\ \hat{\mathbf{x}}_{\mathbf{0}} \quad \dots \quad \hat{\mathbf{x}}_{\mathbf{T}-1} \in \mathcal{X} & \end{array}$$

subject to

$$\begin{aligned} \left| \hat{\mathbf{r}}_{\mathbf{t}} - r^{(u_t), k_t} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_{\mathbf{t}} - x^{(u_t), k_t} \right\|^2, \forall (t, k_t) \in \{0, \dots, T-1\} \times \{1, \dots, n^{(u_t)}\}, \\ \left\| \hat{\mathbf{x}}_{\mathbf{t}+1} - y^{(u_t), k_t} \right\|^2 &\leq L_f^2 \left\| \hat{\mathbf{x}}_{\mathbf{t}} - x^{(u_t), k_t} \right\|^2, \forall (t, k_t) \in \{0, \dots, T-1\} \times \{1, \dots, n^{(u_t)}\}, \\ |\hat{\mathbf{r}}_{\mathbf{t}} - \hat{\mathbf{r}}_{\mathbf{t}'}|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_{\mathbf{t}} - \hat{\mathbf{x}}_{\mathbf{t}'} \right\|^2, \forall t, t' \in \{0, \dots, T-1 | u_t = u_{t'}\}, \\ \left\| \hat{\mathbf{x}}_{\mathbf{t}+1} - \hat{\mathbf{x}}_{\mathbf{t}'+1} \right\|^2 &\leq L_f^2 \left\| \hat{\mathbf{x}}_{\mathbf{t}} - \hat{\mathbf{x}}_{\mathbf{t}'} \right\|^2, \forall t, t' \in \{0, \dots, T-2 | u_t = u_{t'}\}, \\ \hat{\mathbf{x}}_{\mathbf{0}} &= x_0. \end{aligned}$$

The two-stage case

$(\mathcal{P}_2(\mathcal{F}, L_f, L_\rho, x_0, u_0, u_1)) :$

$$\begin{aligned} & \min && \hat{\mathbf{r}}_0 + \hat{\mathbf{r}}_1, \\ & \hat{\mathbf{r}}_0, \hat{\mathbf{r}}_1 \in \mathbb{R} \\ & \hat{\mathbf{x}}_0, \hat{\mathbf{x}}_1 \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} & \left| \hat{\mathbf{r}}_0 - r^{(u_0), k_0} \right|^2 \leq L_\rho^2 \left\| \hat{\mathbf{x}}_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \left\{ 1, \dots, n^{(u_0)} \right\}, \\ & \left| \hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right|^2 \leq L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2, \forall k_1 \in \left\{ 1, \dots, n^{(u_1)} \right\}, \\ & \left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 \leq L_f^2 \left\| \hat{\mathbf{x}}_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \left\{ 1, \dots, n^{(u_0)} \right\}, \\ & \left| \hat{\mathbf{r}}_0 - \hat{\mathbf{r}}_1 \right|^2 \leq L_\rho^2 \left\| \hat{\mathbf{x}}_0 - \hat{\mathbf{x}}_1 \right\|^2 \text{ if } u_0 = u_1, \\ & \hat{\mathbf{x}}_0 = x_0. \end{aligned}$$

$$\left(\mathcal{P}_2'^{(u_0, u_1)}\right) :$$

$$\begin{aligned} & \min && \hat{\mathbf{r}}_0 \\ & \hat{\mathbf{r}}_0 \in \mathbb{R} \\ & \hat{\mathbf{x}}_0 \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} \left| \hat{\mathbf{r}}_0 - r^{(u_0), k_0} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \left\{ 1, \dots, n^{(u_0)} \right\}, \\ \hat{\mathbf{x}}_0 &= x_0. \end{aligned}$$

$$\left(\mathcal{P}_2''^{(u_0, u_1)}\right) :$$

$$\begin{aligned} & \min && \hat{\mathbf{r}}_1 \\ & \hat{\mathbf{r}}_1 \in \mathbb{R} \\ & \hat{\mathbf{x}}_1 \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} \left| \hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2, \forall k_1 \in \left\{ 1, \dots, n^{(u_1)} \right\}, \\ \left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 &\leq L_f^2 \left\| x_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \left\{ 1, \dots, n^{(u_0)} \right\}. \end{aligned}$$

Decomposition

Theorem *Let $(u_0, u_1) \in \mathcal{U}^2$. If $(\hat{\mathbf{r}}_0^*, \hat{\mathbf{x}}_0^*)$ is an optimal solution to $\left(\mathcal{P}_2'^{(u_0, u_1)}\right)$ and $(\hat{\mathbf{r}}_1^*, \hat{\mathbf{x}}_1^*)$ is an optimal solution to $\left(\mathcal{P}_2''^{(u_0, u_1)}\right)$, then $(\hat{\mathbf{r}}_0^*, \hat{\mathbf{r}}_1^*, \hat{\mathbf{x}}_0^*, \hat{\mathbf{x}}_1^*)$ is an optimal solution to $\left(\mathcal{P}_2^{(u_0, u_1)}\right)$.*

Decomposition

Theorem *Let $(u_0, u_1) \in \mathcal{U}^2$. If $(\hat{\mathbf{r}}_0^*, \hat{\mathbf{x}}_0^*)$ is an optimal solution to $(\mathcal{P}_2'^{(u_0, u_1)})$ and $(\hat{\mathbf{r}}_1^*, \hat{\mathbf{x}}_1^*)$ is an optimal solution to $(\mathcal{P}_2''^{(u_0, u_1)})$, then $(\hat{\mathbf{r}}_0^*, \hat{\mathbf{r}}_1^*, \hat{\mathbf{x}}_0^*, \hat{\mathbf{x}}_1^*)$ is an optimal solution to $(\mathcal{P}_2^{(u_0, u_1)})$.*

Corollary *The solution of the problem $(\mathcal{P}_2'^{(u_0, u_1)})$ is*

$$\hat{\mathbf{r}}_0^* = \max_{k_0 \in \{1, \dots, n^{(u_0)}\}} r^{(u_0), k_0} - L_\rho \left\| x_0 - x^{(u_0), k_0} \right\| .$$

Decomposition

Theorem *Let $(u_0, u_1) \in \mathcal{U}^2$. If $(\hat{\mathbf{r}}_0^*, \hat{\mathbf{x}}_0^*)$ is an optimal solution to $(\mathcal{P}_2'^{(u_0, u_1)})$ and $(\hat{\mathbf{r}}_1^*, \hat{\mathbf{x}}_1^*)$ is an optimal solution to $(\mathcal{P}_2''^{(u_0, u_1)})$, then $(\hat{\mathbf{r}}_0^*, \hat{\mathbf{r}}_1^*, \hat{\mathbf{x}}_0^*, \hat{\mathbf{x}}_1^*)$ is an optimal solution to $(\mathcal{P}_2^{(u_0, u_1)})$.*

Corollary *The solution of the problem $(\mathcal{P}_2'^{(u_0, u_1)})$ is*

$$\hat{\mathbf{r}}_0^* = \max_{k_0 \in \{1, \dots, n^{(u_0)}\}} r^{(u_0), k_0} - L_\rho \left\| x_0 - x^{(u_0), k_0} \right\| .$$

Theorem $(\mathcal{P}_2''^{(u_0, u_1)})$ is NP-hard.

Decomposition

Theorem *Let $(u_0, u_1) \in \mathcal{U}^2$. If $(\hat{\mathbf{r}}_0^*, \hat{\mathbf{x}}_0^*)$ is an optimal solution to $(\mathcal{P}_2'^{(u_0, u_1)})$ and $(\hat{\mathbf{r}}_1^*, \hat{\mathbf{x}}_1^*)$ is an optimal solution to $(\mathcal{P}_2''^{(u_0, u_1)})$, then $(\hat{\mathbf{r}}_0^*, \hat{\mathbf{r}}_1^*, \hat{\mathbf{x}}_0^*, \hat{\mathbf{x}}_1^*)$ is an optimal solution to $(\mathcal{P}_2^{(u_0, u_1)})$.*

Corollary *The solution of the problem $(\mathcal{P}_2'^{(u_0, u_1)})$ is*

$$\hat{\mathbf{r}}_0^* = \max_{k_0 \in \{1, \dots, n^{(u_0)}\}} r^{(u_0), k_0} - L_\rho \left\| x_0 - x^{(u_0), k_0} \right\| .$$

Theorem $(\mathcal{P}_2''^{(u_0, u_1)})$ is NP-hard.

Theorem *The two-stage problem $(\mathcal{P}_2^{(u_0, u_1)})$ and the generalized T -stage problem $(\mathcal{P}_T(\mathcal{F}, L_f, L_\rho, x_0, u_0, \dots, u_{T-1}))$ are NP-hard.*

Decomposition

$(\mathcal{P}_2(\mathcal{F}, L_f, L_\rho, x_0, u_0, u_1)) :$

$$\begin{aligned} & \min && \hat{\mathbf{r}}_0 + \hat{\mathbf{r}}_1, \\ & \hat{\mathbf{r}}_0, \hat{\mathbf{r}}_1 \in \mathbb{R} \\ & \hat{\mathbf{x}}_0, \hat{\mathbf{x}}_1 \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} & \left| \hat{\mathbf{r}}_0 - r^{(u_0), k_0} \right|^2 \leq L_\rho^2 \left\| \hat{\mathbf{x}}_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \left\{ 1, \dots, n^{(u_0)} \right\}, \\ & \left| \hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right|^2 \leq L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2, \forall k_1 \in \left\{ 1, \dots, n^{(u_1)} \right\}, \\ & \left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 \leq L_f^2 \left\| \hat{\mathbf{x}}_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \left\{ 1, \dots, n^{(u_0)} \right\}, \\ & |\hat{\mathbf{r}}_0 - \hat{\mathbf{r}}_1|^2 \leq L_\rho^2 \left\| \hat{\mathbf{x}}_0 - \hat{\mathbf{x}}_1 \right\|^2 \text{ if } u_0 = u_1, \\ & \hat{\mathbf{x}}_0 = x_0. \end{aligned}$$

Decomposition

$(\mathcal{P}_2(\mathcal{F}, L_f, L_\rho, x_0, u_0, u_1)) :$

$$\begin{aligned} \min \quad & \hat{\mathbf{r}}_0 + \hat{\mathbf{r}}_1, \\ \hat{\mathbf{r}}_0, \hat{\mathbf{r}}_1 \in \mathbb{R} \\ \hat{\mathbf{x}}_0, \hat{\mathbf{x}}_1 \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} \left| \hat{\mathbf{r}}_0 - r^{(u_0), k_0} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \{1, \dots, n^{(u_0)}\}, \\ \left| \hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2, \forall k_1 \in \{1, \dots, n^{(u_1)}\}, \\ \left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 &\leq L_f^2 \left\| \hat{\mathbf{x}}_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \{1, \dots, n^{(u_0)}\}, \\ |\hat{\mathbf{r}}_0 - \hat{\mathbf{r}}_1|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_0 - \hat{\mathbf{x}}_1 \right\|^2 \text{ if } u_0 = u_1, \\ \hat{\mathbf{x}}_0 &= x_0. \end{aligned}$$

Equivalent

$(\mathcal{P}_2^{(u_0, u_1)}) :$

$$\begin{aligned} \min \quad & \hat{\mathbf{r}}_0 \\ \hat{\mathbf{r}}_0 \in \mathbb{R} \\ \hat{\mathbf{x}}_0 \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} \left| \hat{\mathbf{r}}_0 - r^{(u_0), k_0} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \{1, \dots, n^{(u_0)}\}, \\ \hat{\mathbf{x}}_0 &= x_0. \end{aligned}$$

$(\mathcal{P}_2^{(u_0, u_1)}) :$

$$\begin{aligned} \min \quad & \hat{\mathbf{r}}_1 \\ \hat{\mathbf{r}}_1 \in \mathbb{R} \\ \hat{\mathbf{x}}_1 \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} \left| \hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2, \forall k_1 \in \{1, \dots, n^{(u_1)}\}, \\ \left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 &\leq L_f^2 \left\| x_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \{1, \dots, n^{(u_0)}\}. \end{aligned}$$

Decomposition

$(\mathcal{P}_2(\mathcal{F}, L_f, L_\rho, x_0, u_0, u_1)) :$

$$\begin{aligned} \min \quad & \hat{\mathbf{r}}_0 + \hat{\mathbf{r}}_1, \\ & \hat{\mathbf{r}}_0, \hat{\mathbf{r}}_1 \in \mathbb{R} \\ & \hat{\mathbf{x}}_0, \hat{\mathbf{x}}_1 \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} \left| \hat{\mathbf{r}}_0 - r^{(u_0), k_0} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \{1, \dots, n^{(u_0)}\}, \\ \left| \hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2, \forall k_1 \in \{1, \dots, n^{(u_1)}\}, \\ \left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 &\leq L_f^2 \left\| \hat{\mathbf{x}}_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \{1, \dots, n^{(u_0)}\}, \\ \left| \hat{\mathbf{r}}_0 - \hat{\mathbf{r}}_1 \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_0 - \hat{\mathbf{x}}_1 \right\|^2 \text{ if } u_0 = u_1, \\ \hat{\mathbf{x}}_0 &= x_0. \end{aligned}$$

Equivalent

$(\mathcal{P}_2^{(u_0, u_1)}) :$

$$\begin{aligned} \min \quad & \hat{\mathbf{r}}_0 \\ & \hat{\mathbf{r}}_0 \in \mathbb{R} \\ & \hat{\mathbf{x}}_0 \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} \left| \hat{\mathbf{r}}_0 - r^{(u_0), k_0} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \{1, \dots, n^{(u_0)}\}, \\ \hat{\mathbf{x}}_0 &= x_0. \end{aligned}$$

Solved
(closed-form)

$(\mathcal{P}_2^{(u_0, u_1)}) :$

$$\begin{aligned} \min \quad & \hat{\mathbf{r}}_1 \\ & \hat{\mathbf{r}}_1 \in \mathbb{R} \\ & \hat{\mathbf{x}}_1 \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} \left| \hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2, \forall k_1 \in \{1, \dots, n^{(u_1)}\}, \\ \left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 &\leq L_f^2 \left\| x_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \{1, \dots, n^{(u_0)}\}, \end{aligned}$$

NP-hard

Decomposition

$(\mathcal{P}_2(\mathcal{F}, L_f, L_\rho, x_0, u_0, u_1)) :$

$$\begin{aligned} \min \quad & \hat{\mathbf{r}}_0 + \hat{\mathbf{r}}_1, \\ & \hat{\mathbf{r}}_0, \hat{\mathbf{r}}_1 \in \mathbb{R} \\ & \hat{\mathbf{x}}_0, \hat{\mathbf{x}}_1 \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} \left| \hat{\mathbf{r}}_0 - r^{(u_0), k_0} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \{1, \dots, n^{(u_0)}\}, \\ \left| \hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2, \forall k_1 \in \{1, \dots, n^{(u_1)}\}, \\ \left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 &\leq L_f^2 \left\| \hat{\mathbf{x}}_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \{1, \dots, n^{(u_0)}\}, \\ \left| \hat{\mathbf{r}}_0 - \hat{\mathbf{r}}_1 \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_0 - \hat{\mathbf{x}}_1 \right\|^2 \text{ if } u_0 = u_1, \\ \hat{\mathbf{x}}_0 &= x_0. \end{aligned}$$

NP-hard

Equivalent

$(\mathcal{P}_2^{(u_0, u_1)}) :$

$$\begin{aligned} \min \quad & \hat{\mathbf{r}}_0 \\ & \hat{\mathbf{r}}_0 \in \mathbb{R} \\ & \hat{\mathbf{x}}_0 \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} \left| \hat{\mathbf{r}}_0 - r^{(u_0), k_0} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \{1, \dots, n^{(u_0)}\}, \\ \hat{\mathbf{x}}_0 &= x_0. \end{aligned}$$

Solved
(closed-form)

$(\mathcal{P}_2^{(u_0, u_1)}) :$

$$\begin{aligned} \min \quad & \hat{\mathbf{r}}_1 \\ & \hat{\mathbf{r}}_1 \in \mathbb{R} \\ & \hat{\mathbf{x}}_1 \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} \left| \hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2, \forall k_1 \in \{1, \dots, n^{(u_1)}\}, \\ \left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 &\leq L_f^2 \left\| x_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \{1, \dots, n^{(u_0)}\}, \end{aligned}$$

NP-hard

Decomposition

$(\mathcal{P}_2(\mathcal{F}, L_f, L_\rho, x_0, u_0, u_1)) :$

$$\begin{aligned} \min \quad & \hat{\mathbf{r}}_0 + \hat{\mathbf{r}}_1, \\ & \hat{\mathbf{r}}_0, \hat{\mathbf{r}}_1 \in \mathbb{R} \\ & \hat{\mathbf{x}}_0, \hat{\mathbf{x}}_1 \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} \left\| \hat{\mathbf{r}}_0 - r^{(u_0), k_0} \right\|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \{1, \dots, n^{(u_0)}\}, \\ \left\| \hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right\|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2, \forall k_1 \in \{1, \dots, n^{(u_1)}\}, \\ \left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 &\leq L_f^2 \left\| \hat{\mathbf{x}}_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \{1, \dots, n^{(u_0)}\}, \\ \left\| \hat{\mathbf{r}}_0 - \hat{\mathbf{r}}_1 \right\|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_0 - \hat{\mathbf{x}}_1 \right\|^2 \text{ if } u_0 = u_1, \\ \hat{\mathbf{x}}_0 &= x_0. \end{aligned}$$

NP-hard

$(\mathcal{P}_T(\mathcal{F}, L_f, L_\rho, x_0, u_0, \dots, u_{T-1})) :$

$$\begin{aligned} \min \quad & \sum_{t=0}^{T-1} \hat{\mathbf{r}}_t, \\ & \hat{\mathbf{r}}_0 \dots \hat{\mathbf{r}}_{T-1} \in \mathbb{R} \\ & \hat{\mathbf{x}}_0 \dots \hat{\mathbf{x}}_{T-1} \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} \left\| \hat{\mathbf{r}}_t - r^{(u_t), k_t} \right\|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_t - x^{(u_t), k_t} \right\|^2, \forall (t, k_t) \in \{0, \dots, T-1\} \times \{1, \dots, n^{(u_t)}\}, \\ \left\| \hat{\mathbf{x}}_{t+1} - y^{(u_t), k_t} \right\|^2 &\leq L_f^2 \left\| \hat{\mathbf{x}}_t - x^{(u_t), k_t} \right\|^2, \forall (t, k_t) \in \{0, \dots, T-1\} \times \{1, \dots, n^{(u_t)}\}, \\ \left\| \hat{\mathbf{r}}_t - \hat{\mathbf{r}}_{t'} \right\|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_t - \hat{\mathbf{x}}_{t'} \right\|^2, \forall t, t' \in \{0, \dots, T-1 \mid u_t = u_{t'}\}, \\ \left\| \hat{\mathbf{x}}_{t+1} - \hat{\mathbf{x}}_{t'+1} \right\|^2 &\leq L_f^2 \left\| \hat{\mathbf{x}}_t - \hat{\mathbf{x}}_{t'} \right\|^2, \forall t, t' \in \{0, \dots, T-2 \mid u_t = u_{t'}\}, \\ \hat{\mathbf{x}}_0 &= x_0. \end{aligned}$$

NP-hard

Equivalent

$(\mathcal{P}_2^{(u_0, u_1)}) :$

$$\begin{aligned} \min \quad & \hat{\mathbf{r}}_0 \\ & \hat{\mathbf{r}}_0 \in \mathbb{R} \\ & \hat{\mathbf{x}}_0 \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} \left\| \hat{\mathbf{r}}_0 - r^{(u_0), k_0} \right\|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \{1, \dots, n^{(u_0)}\}, \\ \hat{\mathbf{x}}_0 &= x_0. \end{aligned}$$

Solved
(closed-form)

$(\mathcal{P}_2^{(u_0, u_1)}) :$

$$\begin{aligned} \min \quad & \hat{\mathbf{r}}_1 \\ & \hat{\mathbf{r}}_1 \in \mathbb{R} \\ & \hat{\mathbf{x}}_1 \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} \left\| \hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right\|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2, \forall k_1 \in \{1, \dots, n^{(u_1)}\}, \\ \left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 &\leq L_f^2 \left\| x_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \{1, \dots, n^{(u_0)}\}, \end{aligned}$$

NP-hard

$(\mathcal{P}_2''^{(u_0, u_1)}) :$

$$\begin{array}{ll} \min & \hat{\mathbf{r}}_1 \\ \hat{\mathbf{r}}_1 \in \mathbb{R} & \\ \hat{\mathbf{x}}_1 \in \mathcal{X} & \end{array}$$

subject to

$$\begin{aligned} \left\| \hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right\|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2, \forall k_1 \in \{1, \dots, n^{(u_1)}\} \\ \left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 &\leq L_f^2 \left\| x_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \{1, \dots, n^{(u_0)}\} \end{aligned}$$

NP-hard

Building relaxation schemes

- We focus on the following (NP-hard) problem:

$$\left(\mathcal{P}_2''^{(u_0, u_1)}\right) :$$

$$\begin{array}{ll} \min & \hat{\mathbf{r}}_1 \\ & \hat{\mathbf{r}}_1 \in \mathbb{R} \\ & \hat{\mathbf{x}}_1 \in \mathcal{X} \end{array}$$

subject to

$$\begin{aligned} \left| \hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2, \forall k_1 \in \left\{ 1, \dots, n^{(u_1)} \right\}, \\ \left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 &\leq L_f^2 \left\| x_0 - x^{(u_0), k_0} \right\|^2, \forall k_0 \in \left\{ 1, \dots, n^{(u_0)} \right\}. \end{aligned}$$

- We look for relaxation schemes that preserve the nature of min max generalization problem, i.e. offering performance guarantees
- We thus build relaxation schemes providing lower bounds on the return of the sequence of actions

Trust-region relaxation scheme

- We keep one constraint of each type:

$$\left(\mathcal{P}_{TR}''^{(u_0, u_1)}(k_0, k_1) \right) :$$

$$\begin{aligned} & \min && \hat{\mathbf{r}}_1 \\ & \hat{\mathbf{r}}_1 \in \mathbb{R} \\ & \hat{\mathbf{x}}_1 \in \mathcal{X} \end{aligned}$$

subject to

$$\begin{aligned} \left| \hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2, \\ \left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 &\leq L_f^2 \left\| x_0 - x^{(u_0), k_0} \right\|^2. \end{aligned}$$

Trust-region relaxation scheme

Theorem *Let us denote by $B_{TR}''^{(u_0, u_1), k_0, k_1}(\mathcal{F})$ the bound given by the resolution of $(\mathcal{P}_{TR}''^{(u_0, u_1)}(k_0, k_1))$. We have:*

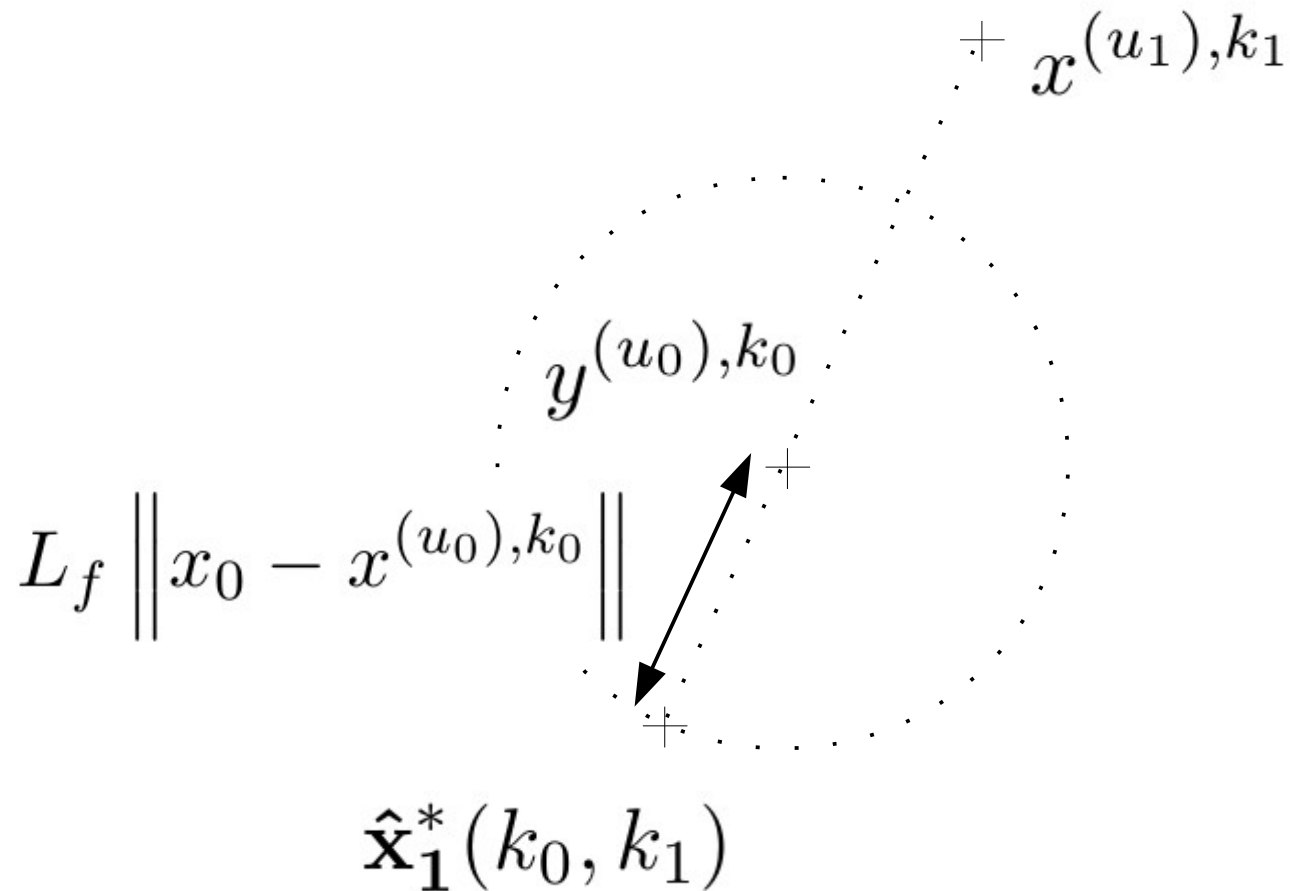
$$B_{TR}''^{(u_0, u_1), k_0, k_1}(\mathcal{F}) = r^{(u_1), k_1} - L_\rho \left\| \hat{\mathbf{x}}_1^*(k_0, k_1) - x^{(u_1), k_1} \right\|,$$

where

$$\hat{\mathbf{x}}_1^*(k_0, k_1) \doteq y^{(u_0), k_0} + L_f \frac{\|x_0 - x^{(u_0), k_0}\|}{\|y^{(u_0), k_0} - x^{(u_1), k_1}\|} \left(y^{(u_0), k_0} - x^{(u_1), k_1} \right) \text{ if } y^{(u_0), k_0} \neq x^{(u_1), k_1}$$

and, if $y^{(u_0), k_0} = x^{(u_1), k_1}$, $\hat{\mathbf{x}}_1^*(k_0, k_1)$ can be any point of the sphere centered in $y^{(u_0), k_0} = x^{(u_1), k_1}$ with radius $L_f \|x_0 - x^{(u_0), k_0}\|$.

Trust-region relaxation scheme



Trust-region relaxation scheme

Theorem *Let us denote by $B_{TR}''^{(u_0, u_1), k_0, k_1}(\mathcal{F})$ the bound given by the resolution of $(\mathcal{P}_{TR}''^{(u_0, u_1)}(k_0, k_1))$. We have:*

$$B_{TR}''^{(u_0, u_1), k_0, k_1}(\mathcal{F}) = r^{(u_1), k_1} - L_\rho \left\| \hat{\mathbf{x}}_1^*(k_0, k_1) - x^{(u_1), k_1} \right\|,$$

where

$$\hat{\mathbf{x}}_1^*(k_0, k_1) \doteq y^{(u_0), k_0} + L_f \frac{\|x_0 - x^{(u_0), k_0}\|}{\|y^{(u_0), k_0} - x^{(u_1), k_1}\|} \left(y^{(u_0), k_0} - x^{(u_1), k_1} \right) \text{ if } y^{(u_0), k_0} \neq x^{(u_1), k_1}$$

and, if $y^{(u_0), k_0} = x^{(u_1), k_1}$, $\hat{\mathbf{x}}_1^*(k_0, k_1)$ can be any point of the sphere centered in $y^{(u_0), k_0} = x^{(u_1), k_1}$ with radius $L_f \|x_0 - x^{(u_0), k_0}\|$.

Definition (Trust-region Bound $B_{TR}^{(u_0, u_1)}(\mathcal{F})$).

$$\forall (u_0, u_1) \in \mathcal{U}^2, \quad B_{TR}^{(u_0, u_1)}(\mathcal{F}) \triangleq \hat{\mathbf{r}}_0^* + \max_{\substack{k_1 \in \{1, \dots, n^{(u_1)}\} \\ k_0 \in \{1, \dots, n^{(u_0)}\}}} B_{TR}''^{(u_0, u_1), k_0, k_1}(\mathcal{F}).$$

Lagrangian relaxation

$$\left(\mathcal{P}_{LD}''^{(u_0, u_1)}\right) :$$

$$\begin{array}{ll} \max & \min \quad \hat{\mathbf{r}}_1 \\ \lambda_1, \dots, \lambda_{n^{(u_0)}} \in \mathbb{R}_+ & \hat{\mathbf{r}}_1 \in \mathbb{R} \\ \mu_1, \dots, \mu_{n^{(u_1)}} \in \mathbb{R}_+ & \hat{\mathbf{x}}_1 \in \mathcal{X} \end{array}$$

$$\begin{aligned} & + \sum_{k_1=1}^{n^{(u_1)}} \mu_{k_1} \left(\left(\hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right)^2 - L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2 \right) \\ & + \sum_{k_0=1}^{n^{(u_0)}} \lambda_{k_0} \left(\left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 - L_f^2 \left\| x_0 - x^{(u_0), k_0} \right\|^2 \right) . \end{aligned}$$

Lagrangian relaxation

$$\left(\mathcal{P}_{LD}''^{(u_0, u_1)}\right) :$$

$$\begin{array}{ll} \max & \min \\ \lambda_1, \dots, \lambda_{n(u_0)} \in \mathbb{R}_+ & \hat{\mathbf{r}}_1 \in \mathbb{R} \\ \mu_1, \dots, \mu_{n(u_1)} \in \mathbb{R}_+ & \hat{\mathbf{x}}_1 \in \mathcal{X} \end{array}$$

$$\begin{aligned} & + \sum_{k_1=1}^{n(u_1)} \mu_{k_1} \left(\left(\hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right)^2 - L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2 \right) \\ & + \sum_{k_0=1}^{n(u_0)} \lambda_{k_0} \left(\left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 - L_f^2 \left\| x_0 - x^{(u_0), k_0} \right\|^2 \right) . \end{aligned}$$

Theorem $\left(\mathcal{P}_{LD}''^{(u_0, u_1)}\right)$ is a conic quadratic program.

Lagrangian relaxation

$$\left(\mathcal{P}_{LD}''^{(u_0, u_1)}\right) :$$

$$\begin{aligned} & \max_{\substack{\lambda_1, \dots, \lambda_{n(u_0)} \in \mathbb{R}_+ \\ \mu_1, \dots, \mu_{n(u_1)} \in \mathbb{R}_+}} \min_{\substack{\hat{\mathbf{r}}_1 \in \mathbb{R} \\ \hat{\mathbf{x}}_1 \in \mathcal{X}}} \hat{\mathbf{r}}_1 \\ & + \sum_{k_1=1}^{n(u_1)} \mu_{k_1} \left(\left(\hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right)^2 - L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2 \right) \\ & + \sum_{k_0=1}^{n(u_0)} \lambda_{k_0} \left(\left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 - L_f^2 \left\| x_0 - x^{(u_0), k_0} \right\|^2 \right) . \end{aligned}$$

Theorem $\left(\mathcal{P}_{LD}''^{(u_0, u_1)}\right)$ is a conic quadratic program.

Definition (Lagrangian Relaxation Bound $B_{LD}^{(u_0, u_1)}(\mathcal{F})$).

$$\forall (u_0, u_1) \in \mathcal{U}^2, \quad B_{LD}^{(u_0, u_1)}(\mathcal{F}) \triangleq \hat{\mathbf{r}}_0^* + B_{LD}''^{(u_0, u_1)}(\mathcal{F})$$

Menu

Introduction

I Direct approach

- CGRL algorithm

II Reformulation of the original problem

- 2 relaxations schemes in the two-stage case

III Comparison of the 3 proposed solutions

Conclusions and future work

Direct approach vs Trust-region relaxation

Definition (CGRL Bound $B_{CGRL}^{(u_0, u_1)}(\mathcal{F})$).

$\forall (u_0, u_1) \in \mathcal{U}^2$,

$$B_{CGRL}^{(u_0, u_1)}(\mathcal{F}) \triangleq \max_{\substack{k_1 \in \{1, \dots, n^{(u_1)}\} \\ k_0 \in \{1, \dots, n^{(u_0)}\}}} r^{(u_0), k_0} - L_\rho(1 + L_f) \left\| x^{(u_0), k_0} - x_0 \right\| \\ + r^{(u_1), k_1} - L_\rho \left\| y^{(u_0), k_0} - x^{(u_1), k_1} \right\|.$$

Direct approach vs Trust-region relaxation

Definition (CGRL Bound $B_{CGRL}^{(u_0, u_1)}(\mathcal{F})$).

$\forall (u_0, u_1) \in \mathcal{U}^2$,

$$B_{CGRL}^{(u_0, u_1)}(\mathcal{F}) \triangleq \max_{\substack{k_1 \in \{1, \dots, n^{(u_1)}\} \\ k_0 \in \{1, \dots, n^{(u_0)}\}}} r^{(u_0), k_0} - L_\rho(1 + L_f) \left\| x^{(u_0), k_0} - x_0 \right\| \\ + r^{(u_1), k_1} - L_\rho \left\| y^{(u_0), k_0} - x^{(u_1), k_1} \right\| .$$

Theorem

$$\forall (u_0, u_1) \in \mathcal{U}^2, \quad B_{CGRL}^{(u_0, u_1)}(\mathcal{F}) \leq B_{TR}^{(u_0, u_1)}(\mathcal{F}) .$$

Trust-region vs Lagrangian relaxation

Lemma *Let $(u_0, u_1) \in \mathcal{U}^2$ and $(k_0, k_1) \in \{1, \dots, n^{(u_0)}\} \times \{1, \dots, n^{(u_1)}\}$. Consider again the problem $\left(\mathcal{P}_{TR}''^{(u_0, u_1)}(k_0, k_1)\right)$ where all constraints are dropped except the two defined by (k_0, k_1) :*

$$\begin{aligned} \left(\mathcal{P}_{TR}''^{(u_0, u_1)}(k_0, k_1)\right) : \quad & \min_{\substack{\hat{\mathbf{r}}_1 \in \mathbb{R} \\ \hat{\mathbf{x}}_1 \in \mathcal{X}}} \quad \hat{\mathbf{r}}_1 \\ \text{subject to} \quad & \left| \hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right|^2 \leq L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2 \\ & \left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 \leq L_f^2 \left\| x_0 - x^{(u_0), k_0} \right\|^2 . \end{aligned}$$

Then, the Lagrangian relaxation of $\left(\mathcal{P}_{TR}''^{(u_0, u_1)}(k_0, k_1)\right)$ leads to a bound denoted by $B_{LD}''^{(u_0, u_1), k_0, k_1}(\mathcal{F})$ which is equal to the Trust-region bound $B_{TR}''^{(u_0, u_1), k_0, k_1}(\mathcal{F})$, i.e.

$$B_{LD}''^{(u_0, u_1), k_0, k_1}(\mathcal{F}) = B_{TR}''^{(u_0, u_1), k_0, k_1}(\mathcal{F}) .$$

Trust-region vs Lagrangian relaxation

Lemma Let $(u_0, u_1) \in \mathcal{U}^2$ and $(k_0, k_1) \in \{1, \dots, n^{(u_0)}\} \times \{1, \dots, n^{(u_1)}\}$. Consider again the problem $\left(\mathcal{P}_{TR}''^{(u_0, u_1)}(k_0, k_1)\right)$ where all constraints are dropped except the two defined by (k_0, k_1) :

$$\begin{aligned} \left(\mathcal{P}_{TR}''^{(u_0, u_1)}(k_0, k_1)\right) : \quad & \min_{\substack{\hat{\mathbf{r}}_1 \in \mathbb{R} \\ \hat{\mathbf{x}}_1 \in \mathcal{X}}} \quad \hat{\mathbf{r}}_1 \\ \text{subject to} \quad & \left\| \hat{\mathbf{r}}_1 - r^{(u_1), k_1} \right\|^2 \leq L_\rho^2 \left\| \hat{\mathbf{x}}_1 - x^{(u_1), k_1} \right\|^2 \\ & \left\| \hat{\mathbf{x}}_1 - y^{(u_0), k_0} \right\|^2 \leq L_f^2 \left\| x_0 - x^{(u_0), k_0} \right\|^2 . \end{aligned}$$

Then, the Lagrangian relaxation of $\left(\mathcal{P}_{TR}''^{(u_0, u_1)}(k_0, k_1)\right)$ leads to a bound denoted by $B_{LD}''^{(u_0, u_1), k_0, k_1}(\mathcal{F})$ which is equal to the Trust-region bound $B_{TR}''^{(u_0, u_1), k_0, k_1}(\mathcal{F})$, i.e.

$$B_{LD}''^{(u_0, u_1), k_0, k_1}(\mathcal{F}) = B_{TR}''^{(u_0, u_1), k_0, k_1}(\mathcal{F}) .$$

Theorem

$$\forall (u_0, u_1) \in \mathcal{U}^2, \quad B_{TR}^{(u_0, u_1)}(\mathcal{F}) \leq B_{LD}^{(u_0, u_1)}(\mathcal{F}).$$

Synthesis

Theorem $\forall (u_0, u_1) \in \mathcal{U}^2,$

$$B_{CGRL}^{(u_0, u_1)}(\mathcal{F}) \leq B_{TR}^{(u_0, u_1)}(\mathcal{F}) \leq B_{LD}^{(u_0, u_1)}(\mathcal{F}) \leq L_2^{(u_0, u_1)}(\mathcal{F}) \leq J_2^{(u_0, u_1)} .$$

Illustration

- Dynamics: $\forall (x, u) \in \mathcal{X} \times \mathcal{U}, \quad f(x, u) = x + 3.1416 \times u \times 1_d$

- Reward function: $\forall (x, u) \in \mathcal{X} \times \mathcal{U}, \quad \rho(x, u) = \sum_{i=1}^d x(i)$

- Initial state: $x_0 = 0.5772 \times 1_d$

- Action space: $\mathcal{U} = \{0, 0.1\}$

- Grid generation:

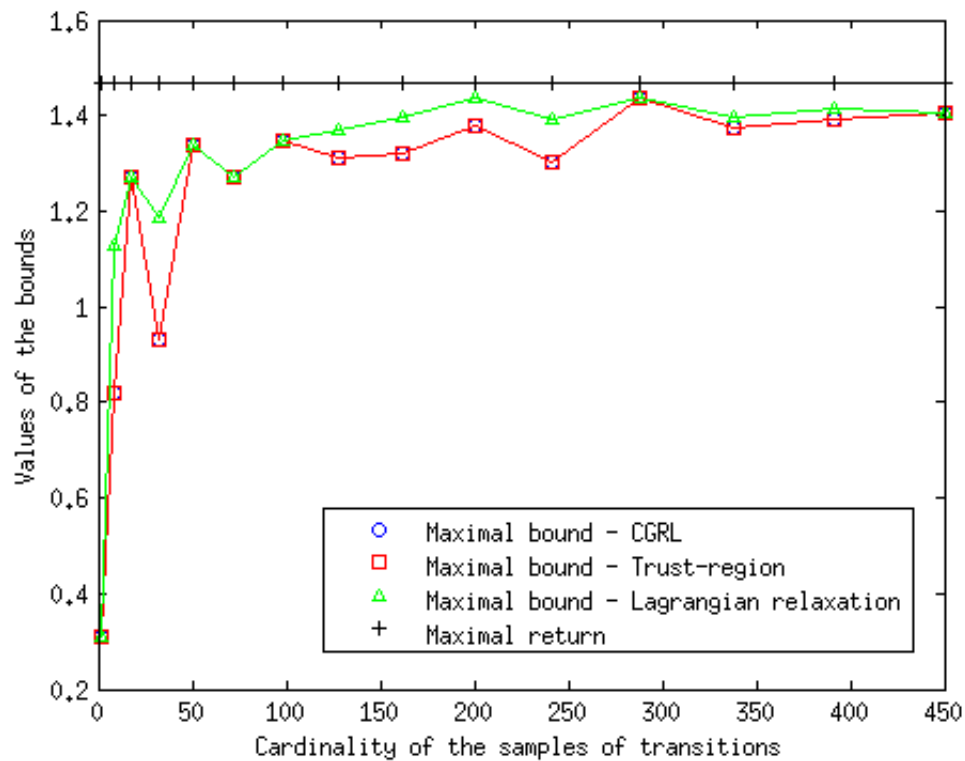
$$\forall u \in \mathcal{U}, \mathcal{F}_{c_i}^{(u)} = \left\{ \left(\left[\frac{i_1}{i}; \frac{i_2}{i} \right], u, \rho \left(\left[\frac{i_1}{i}; \frac{i_2}{i} \right], u \right), f \left(\left[\frac{i_1}{i}; \frac{i_2}{i} \right], u \right) \right) \mid (i_1, i_2) \in \{1, \dots, i\}^2 \right\}$$

- 100 batch collections of transitions drawn uniformly randomly

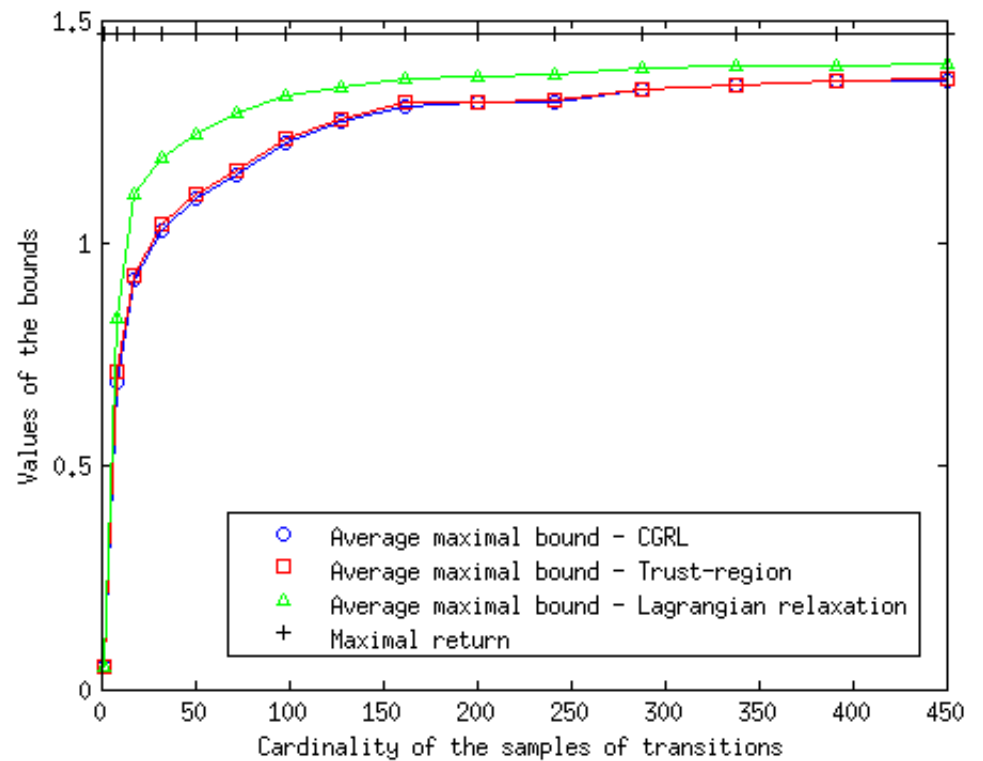
Illustration

Maximal bounds

Grid



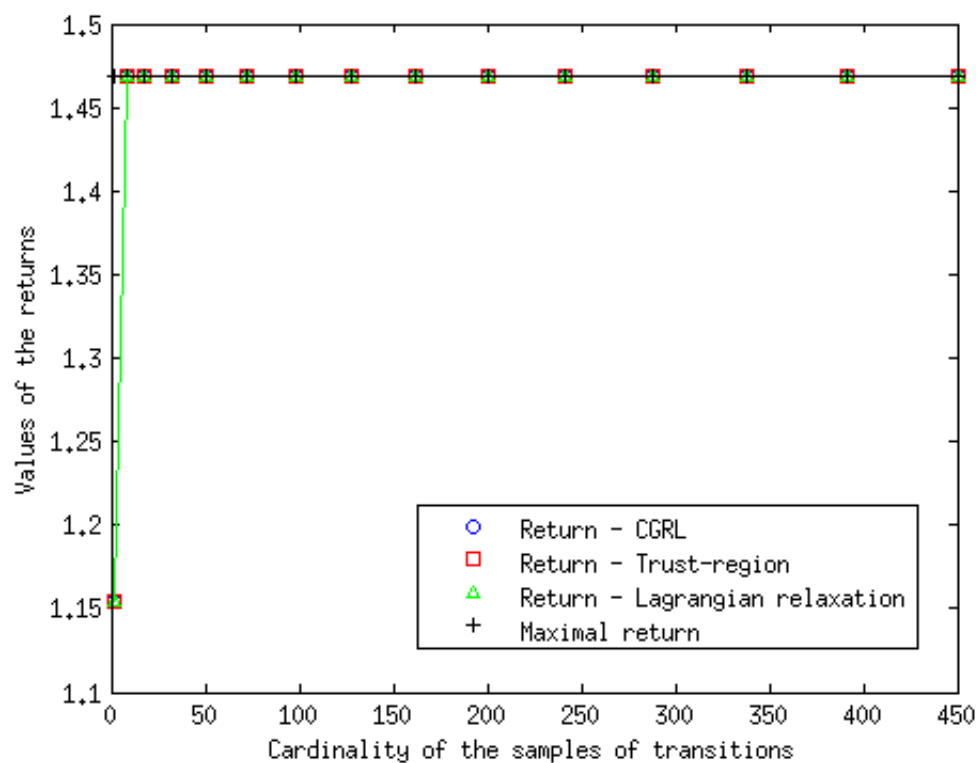
Uniform sampling



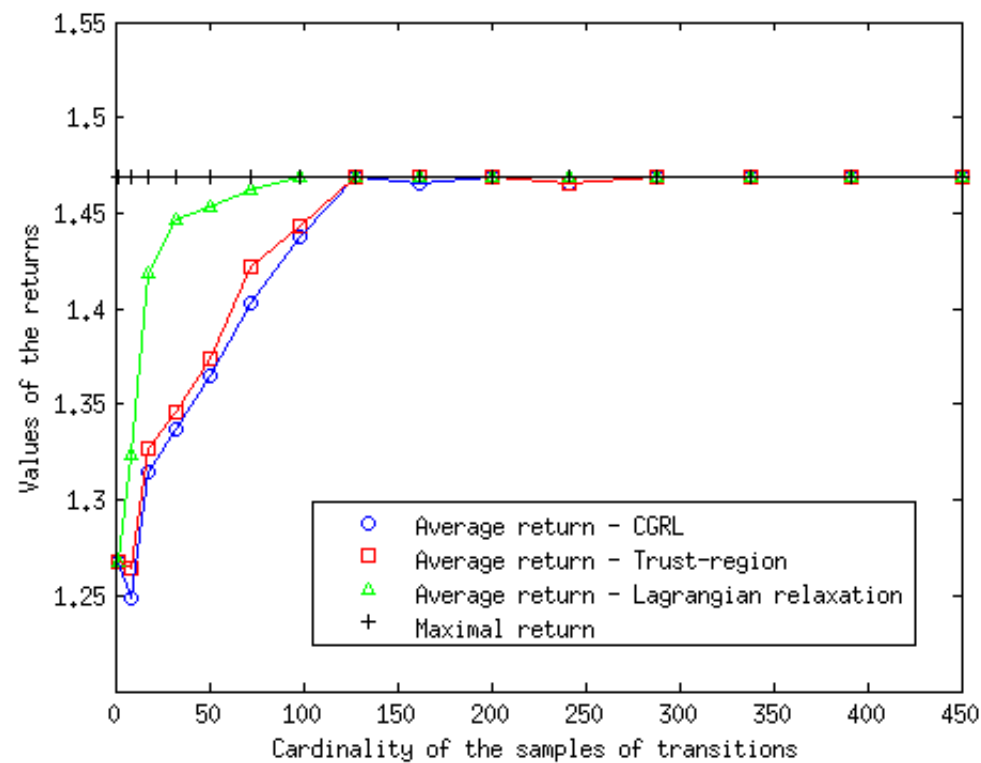
Illustration

Returns

Grid



Uniform sampling



Menu

Introduction

I Direct approach

- CGRL algorithm

II Reformulation of the original problem

- 2 relaxations schemes in the two-stage case

III Comparison of the 3 proposed solutions

Conclusions and future work

(non) Conclusion

$(\mathcal{P}_T(\mathcal{F}, L_f, L_\rho, x_0, u_0, \dots, u_{T-1})) :$

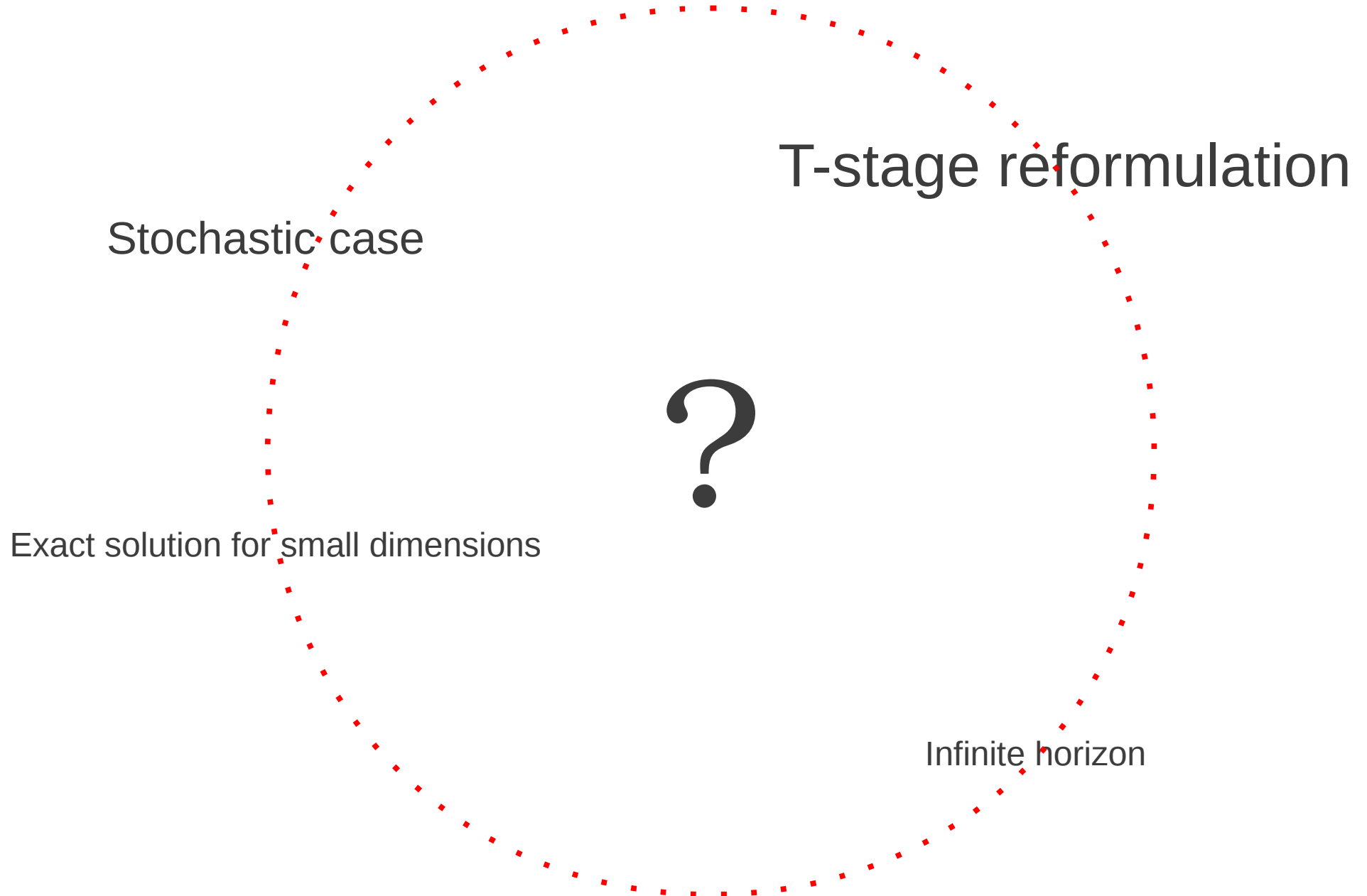
$$\begin{array}{ll} \min & \sum_{t=0}^{T-1} \hat{\mathbf{r}}_{\mathbf{t}}, \\ \hat{\mathbf{r}}_{\mathbf{0}} \quad \dots \quad \hat{\mathbf{r}}_{\mathbf{T}-1} & \in \mathbb{R} \\ \hat{\mathbf{x}}_{\mathbf{0}} \quad \dots \quad \hat{\mathbf{x}}_{\mathbf{T}-1} & \in \mathcal{X} \end{array}$$

s.t.

$$\begin{aligned} \left| \hat{\mathbf{r}}_{\mathbf{t}} - r^{(u_t), k_t} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_{\mathbf{t}} - x^{(u_t), k_t} \right\|^2, \forall (t, k_t) \in \{0, \dots, T-1\} \times \{1, \dots, n^{(u_t)}\}, \\ \left\| \hat{\mathbf{x}}_{\mathbf{t}+1} - y^{(u_t), k_t} \right\|^2 &\leq L_f^2 \left\| \hat{\mathbf{x}}_{\mathbf{t}} - x^{(u_t), k_t} \right\|^2, \forall (t, k_t) \in \{0, \dots, T-1\} \times \{1, \dots, n^{(u_t)}\}, \\ \left| \hat{\mathbf{r}}_{\mathbf{t}} - \hat{\mathbf{r}}_{\mathbf{t}'} \right|^2 &\leq L_\rho^2 \left\| \hat{\mathbf{x}}_{\mathbf{t}} - \hat{\mathbf{x}}_{\mathbf{t}'} \right\|^2, \forall t, t' \in \{0, \dots, T-1 | u_t = u_{t'}\}, \\ \left\| \hat{\mathbf{x}}_{\mathbf{t}+1} - \hat{\mathbf{x}}_{\mathbf{t}'+1} \right\|^2 &\leq L_f^2 \left\| \hat{\mathbf{x}}_{\mathbf{t}} - \hat{\mathbf{x}}_{\mathbf{t}'} \right\|^2, \forall t, t' \in \{0, \dots, T-2 | u_t = u_{t'}\}, \\ \hat{\mathbf{x}}_{\mathbf{0}} &= x_0. \end{aligned}$$

Problem still unsolved

Future work



Associated publications

- **"Min max generalization for deterministic batch mode reinforcement learning: relaxation schemes"**. R. Fonteneau, D. Ernst, B. Boigelot, Q. Louveaux. arXiv:1202.5298v1, 2012.
- **"Towards Min Max Generalization in Reinforcement Learning"**. R. Fonteneau, S.A. Murphy, L. Wehenkel and D. Ernst. Agents and Artificial Intelligence: International Conference, ICAART 2010, Valencia, Spain, January 2010, Revised Selected Papers. Series: Communications in Computed and Information Science (CCIS), Volume 129, pp. 61-77. Editors: J. Filipe, A. Fred, and B.Sharp. Springer, Heidelberg, 2011.
- **"A cautious approach to generalization in reinforcement learning"**. R. Fonteneau, S.A. Murphy, L. Wehenkel and D. Ernst. Proceedings of The International Conference on Agents and Artificial Intelligence (ICAART 2010), 10 pages, Valencia, Spain, January 22-24, 2010
- **"Inferring bounds on the performance of a control policy from a sample of trajectories"**. R. Fonteneau, S.A. Murphy, L. Wehenkel and D. Ernst. Proceedings of The IEEE International Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL 2009), 7 pages, Nashville, Tennessee, USA, 30 March-2 April, 2009.