Model-free Monte Carlo-like Estimation

Looking into proofs…

@R_Fonteneau

April 4th, 2019.

Formalization

Reinforcement learning

System dynamics: $x_{t+1} = f(x_t, u_t, w_t)$

$t \in \{0, \dots, T-1\}$ $x_t \in \mathcal{X} \subset \mathbb{R}^d$ $u_t \in \mathcal{U}$ $w_t \in \mathcal{W}$

$w_t \sim p_{\mathcal{W}}(\cdot)$

Reward function: $r_t = \rho(x_t, u_t, w_t)$

Performance of a policy $h : \{0, \dots, T-1\} \times \mathcal{X} \to \mathcal{U}$

$$J^h(x_0) = \mathbb{E}\left[R^h(x_0, w_0, \dots, w_{T-1})\right]$$

where

$$R^h(x_0, w_0, \dots, w_{T-1}) = \sum_{t=0}^{T-1} \rho(x_t, h(t, x_t), w_t)$$

$$x_{t+1} = f(x_t, h(t, x_t), w_t)$$

Formalization

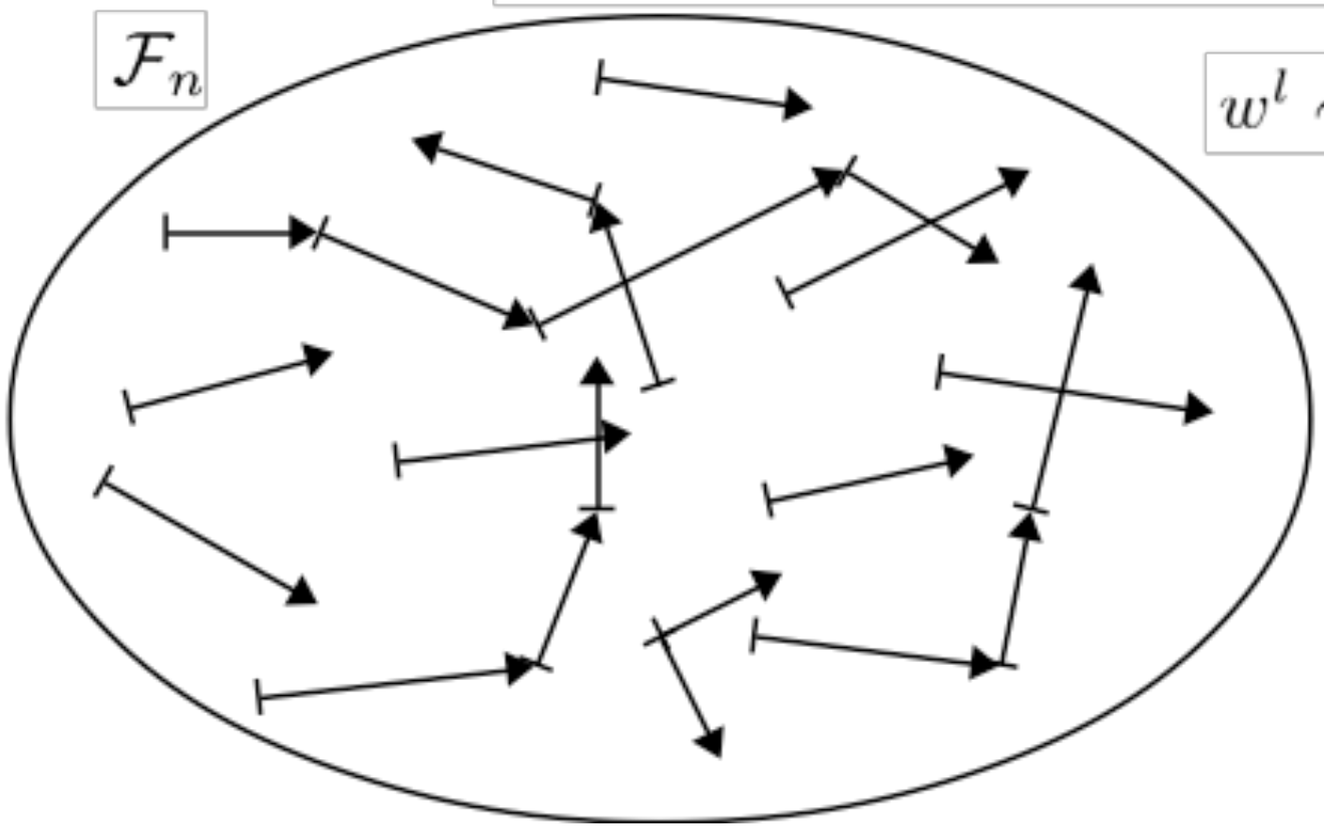
Batch mode reinforcement learning

The system dynamics, reward function and disturbance probability distribution are **unknown**

Instead, we have access to a **sample of one-step system transitions**:

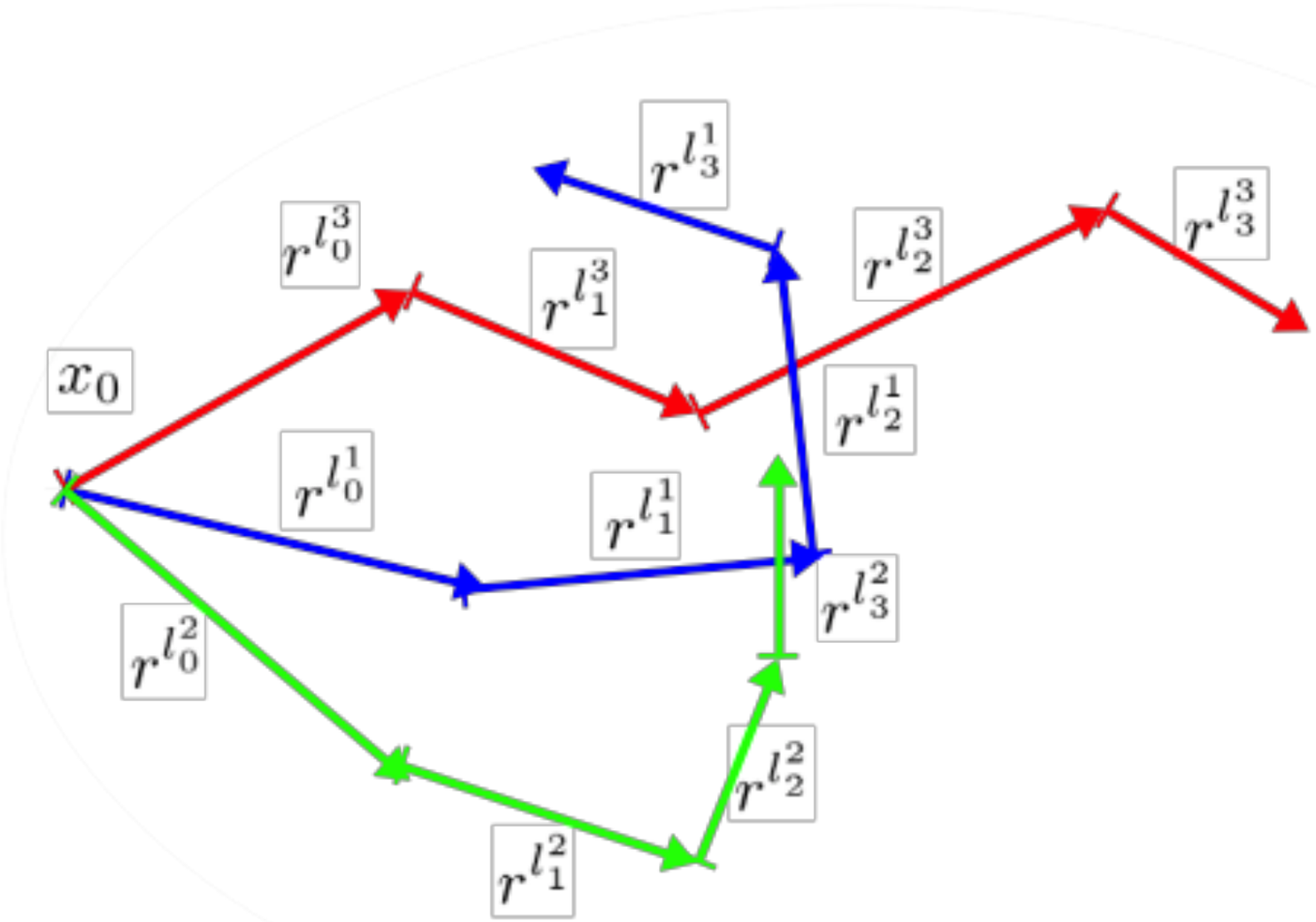
$$\mathcal{F}_n = \left\{ \left(x^l, u^l, r^l, y^l \right) \right\}_{l=1}^{n}$$

$$\forall l \in \{1, \ldots, n\}, \qquad r^l = \rho\left(x^l, u^l, w^l\right)$$
$$y^l = f\left(x^l, u^l, w^l\right)$$



Model-free Monte Carlo Estimation

MODEL OR SIMULATOR REQUIRED!



$$\mathbb{M}_3^h(x_0) = \frac{\left(r^{l_0^1} + r^{l_1^1} + r^{l_2^1} + r^{l_3^1}\right) + \left(r^{l_0^2} + r^{l_1^2} + r^{l_2^2} + r^{l_3^2}\right) + \left(r^{l_0^3} + r^{l_1^3} + r^{l_2^3} + r^{l_3^3}\right)}{3}$$

Model-free Monte Carlo Estimation

If the system dynamics and the reward function were accessible to simulation, then **Monte Carlo (MC) estimation** would allow estimating the performance of *h*

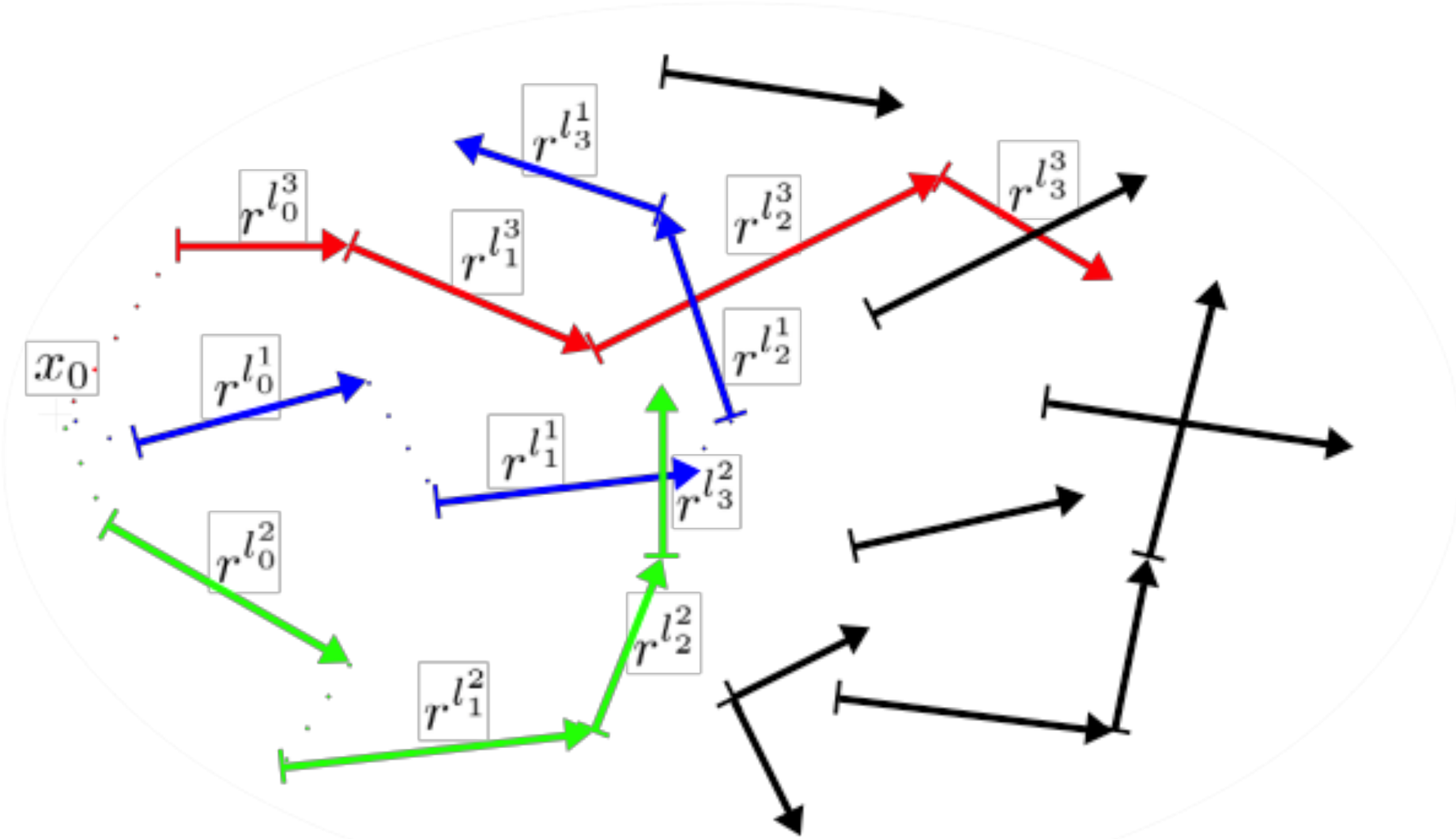
We propose an approach that mimics MC estimation by rebuilding *p* **artificial trajectories** from one-step system transitions

These artificial trajectories are built so as to **minimize the discrepancy (using a distance metric Δ) with a classical MC sample** that could be obtained by simulating the system with the policy *h*; each one step transition is used **at most once**

We average the cumulated returns over the *p* artificial trajectories to obtain the **Model-free Monte Carlo estimator** (MFMC) of the expected return of *h*:

$$\mathfrak{M}_p^h\left(\mathcal{F}_n, x_0\right) = \frac{1}{p}\sum_{i=1}^{p}\sum_{t=0}^{T-1} r^{l_t^i}$$

Model-free Monte Carlo Estimation

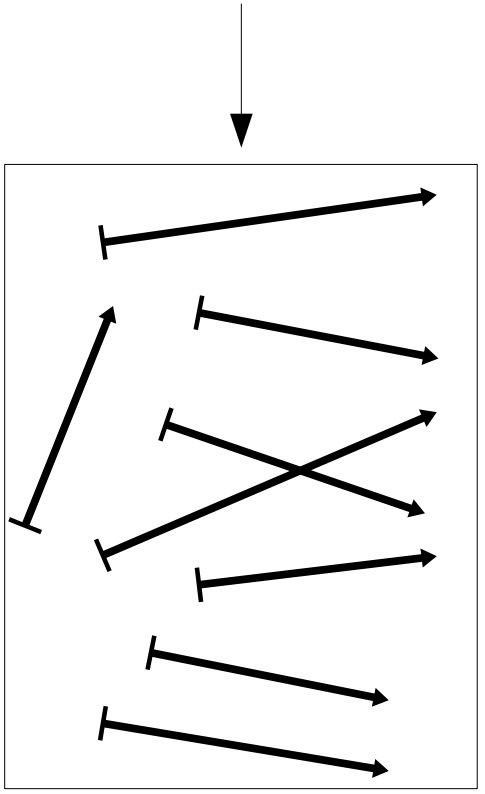


$$\mathfrak{M}_3^h\left(\mathcal{F}_n, x_0\right) = \frac{\left(r^{l_0^1} + r^{l_1^1} + r^{l_2^1} + r^{l_3^1}\right) + \left(r^{l_0^2} + r^{l_1^2} + r^{l_2^2} + r^{l_3^2}\right) + \left(r^{l_0^3} + r^{l_1^3} + r^{l_2^3} + r^{l_3^3}\right)}{3}$$

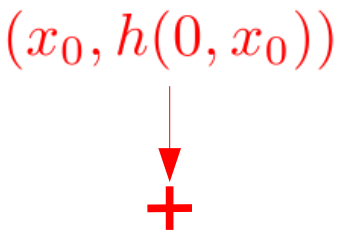
The MFMC algorithm

Example with T = 3, p = 2, n = 8

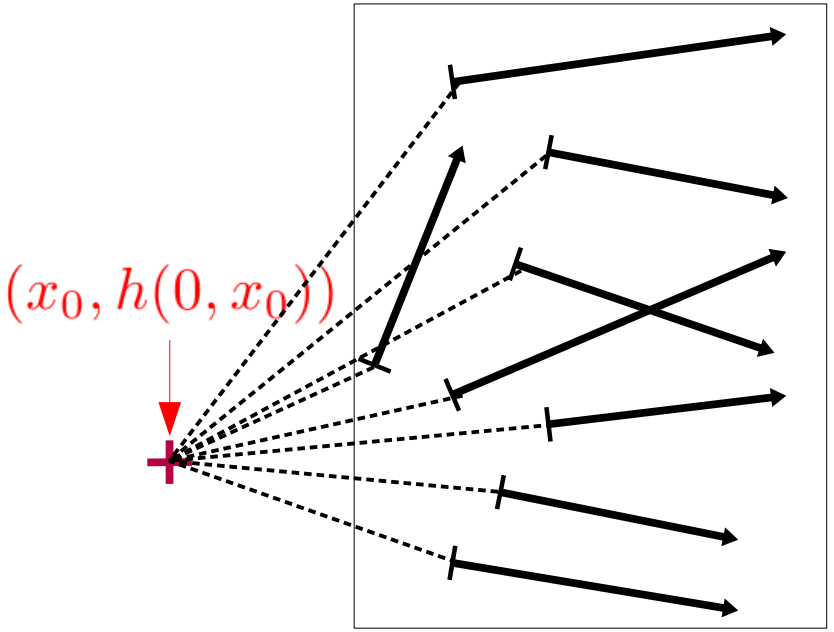
$$\mathcal{F}_n = \left\{ \left(x^l, u^l, r^l, y^l \right) \in \mathcal{X} \times \mathcal{U} \times \mathbb{R} \times \mathcal{X} \right\}_{l=1}^{n}$$



The MFMC algorithm

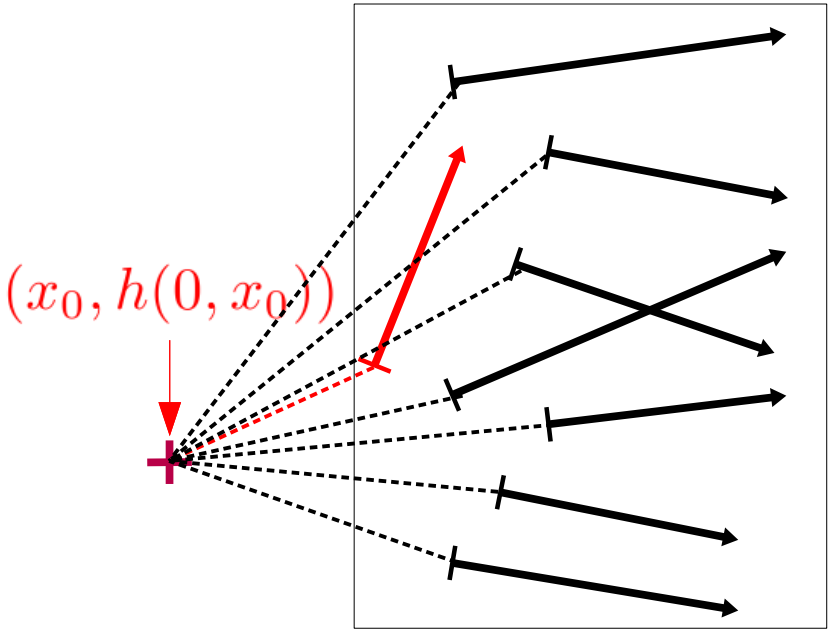


The MFMC algorithm



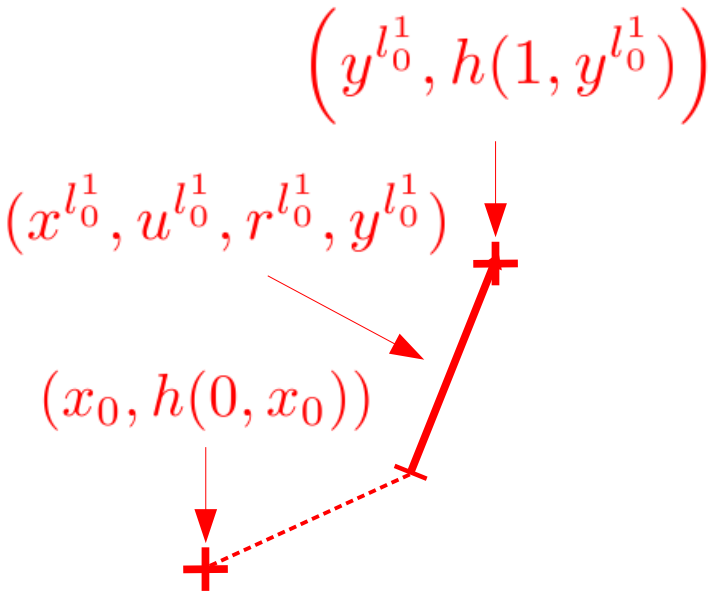
$$\mathcal{G} = \mathcal{F}_n$$

The MFMC algorithm

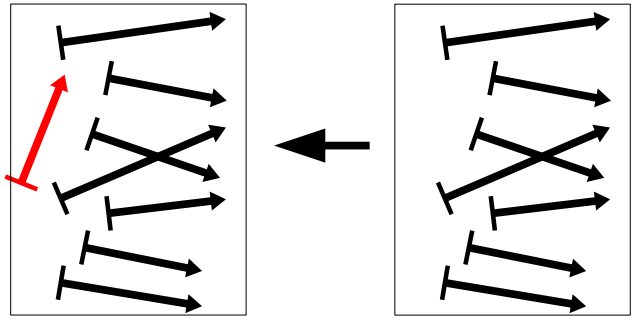


$$\mathcal{G} = \mathcal{F}_n$$

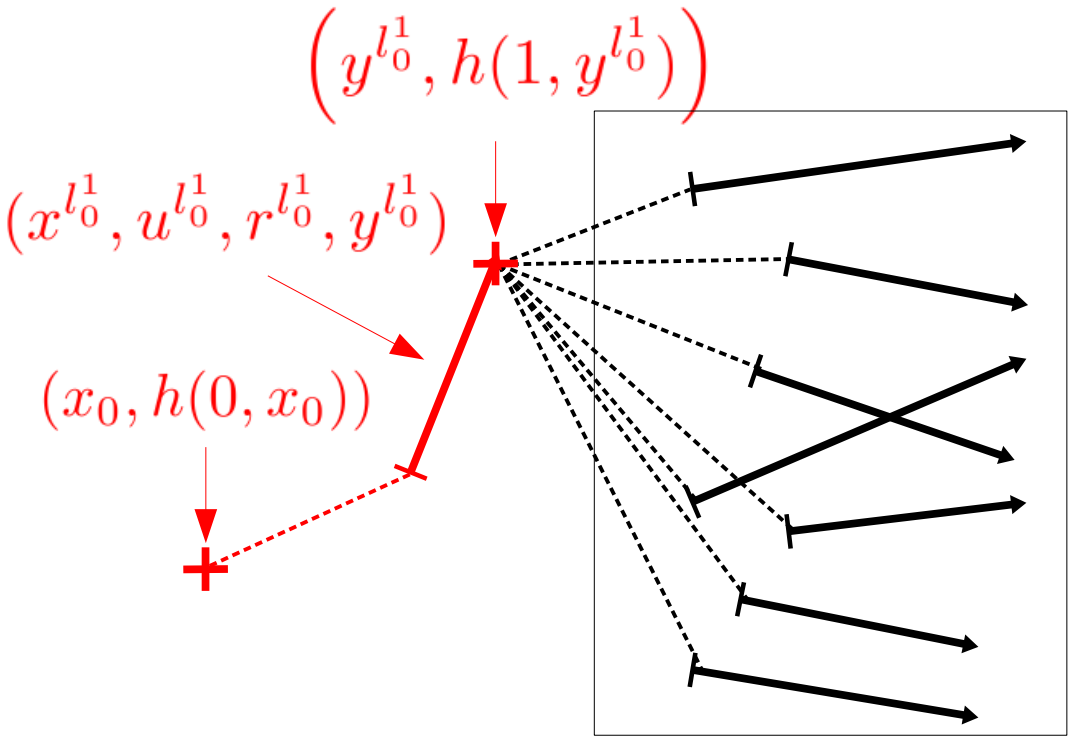
The MFMC algorithm



$$\mathcal{G} = \mathcal{G} \setminus \{(x^{l_0^1}, u^{l_0^1}, r^{l_0^1}, y^{l_0^1})\}$$

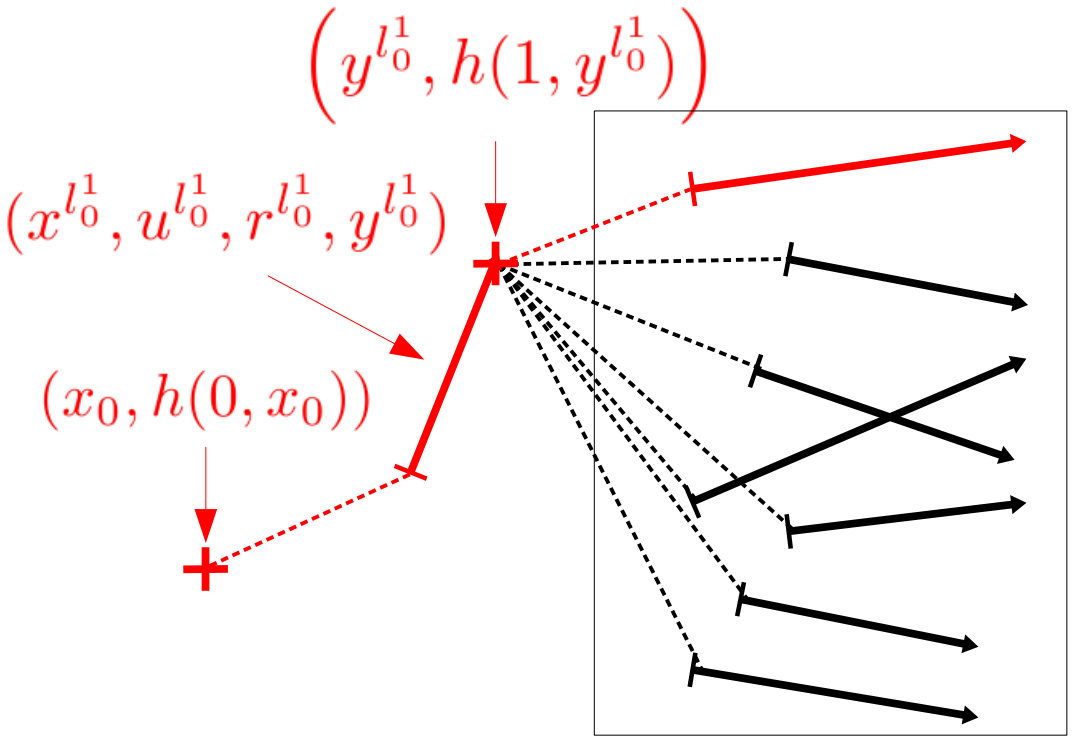


The MFMC algorithm



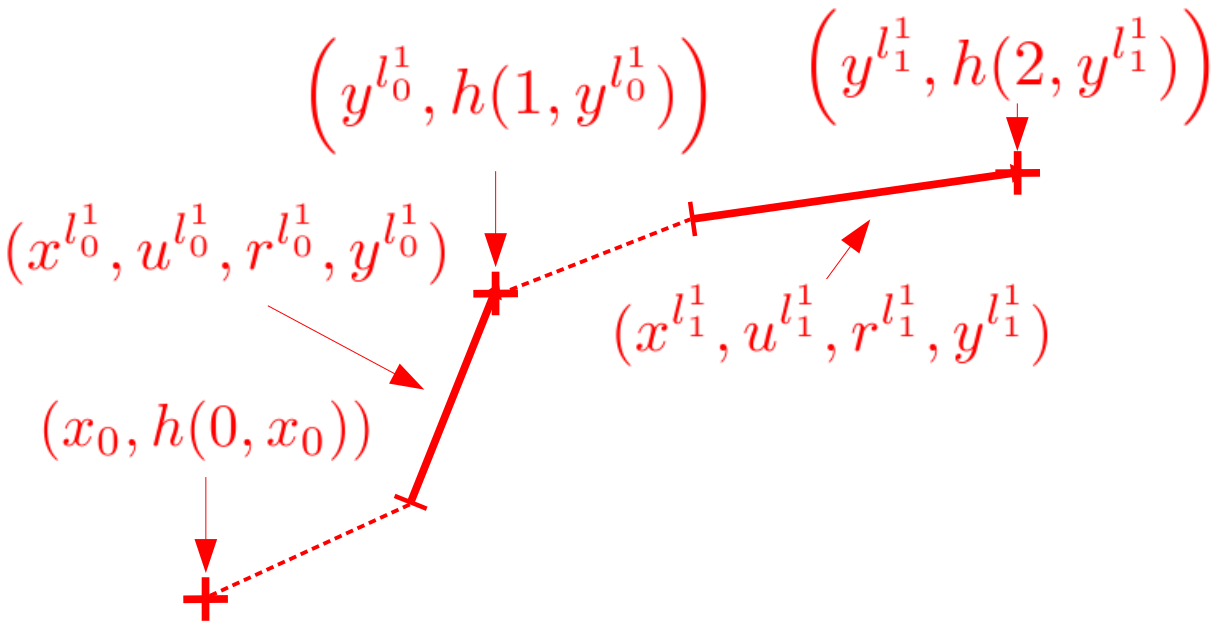
$$\mathcal{G} = \mathcal{G} \setminus \{(x^{l_0^1}, u^{l_0^1}, r^{l_0^1}, y^{l_0^1})\}$$

The MFMC algorithm

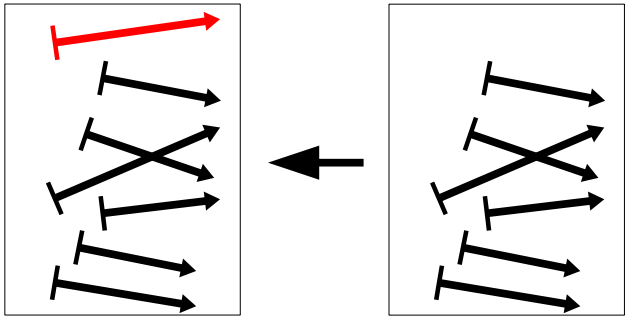


$$\mathcal{G} = \mathcal{G} \setminus \{(x^{l_0^1}, u^{l_0^1}, r^{l_0^1}, y^{l_0^1})\}$$

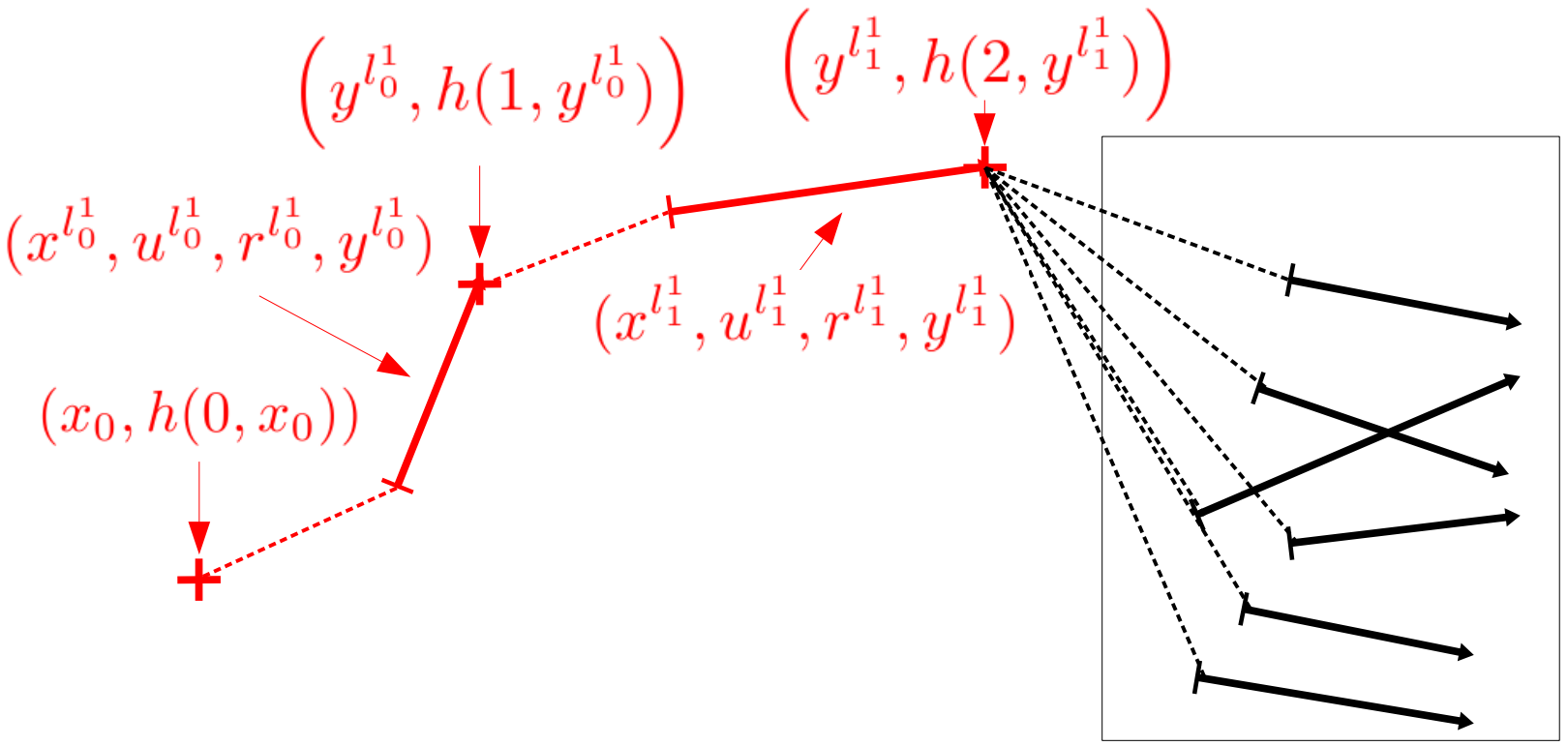
The MFMC algorithm



$$\mathcal{G} = \mathcal{G} \setminus \{(x^{l_1^1}, u^{l_1^1}, r^{l_1^1}, y^{l_1^1})\}$$

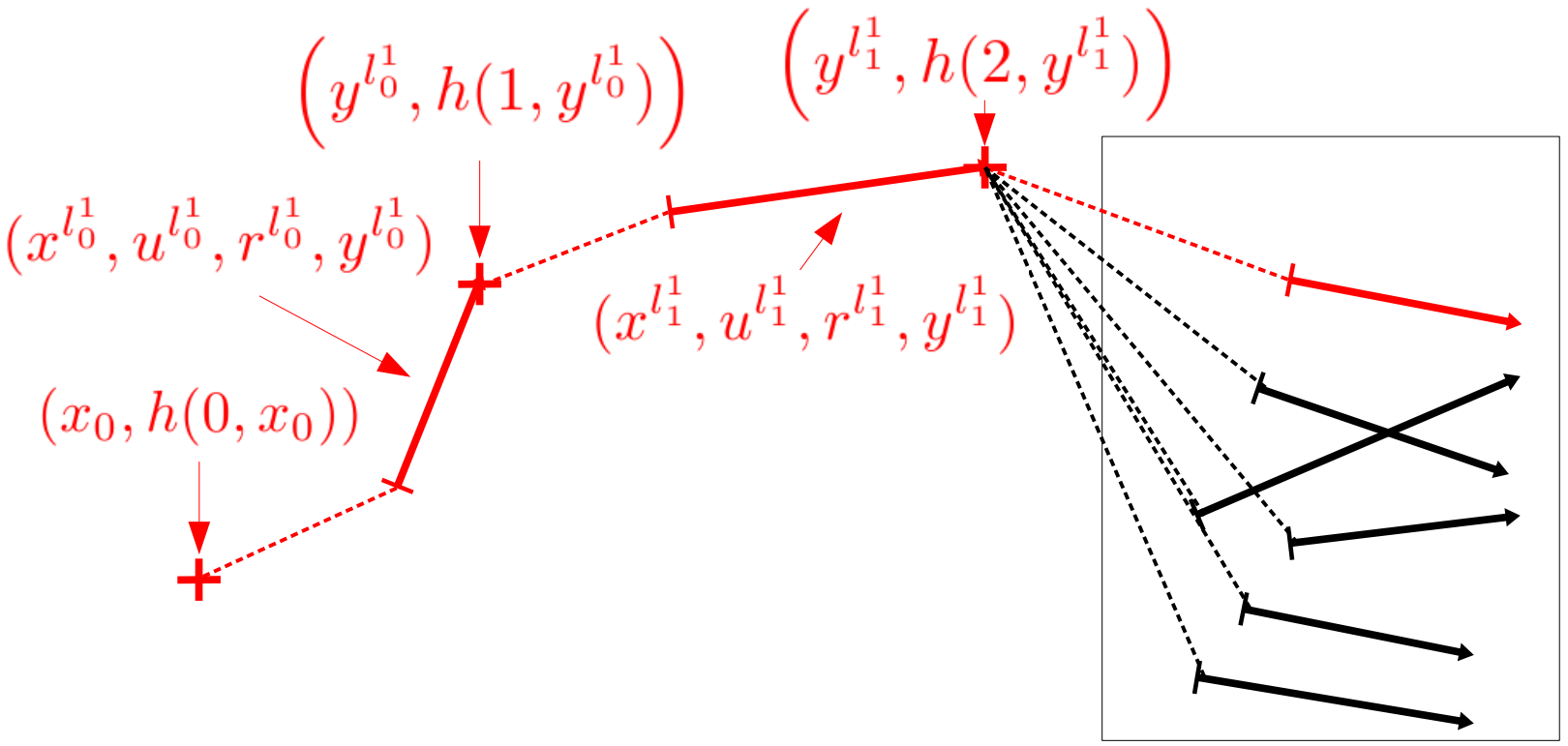


The MFMC algorithm



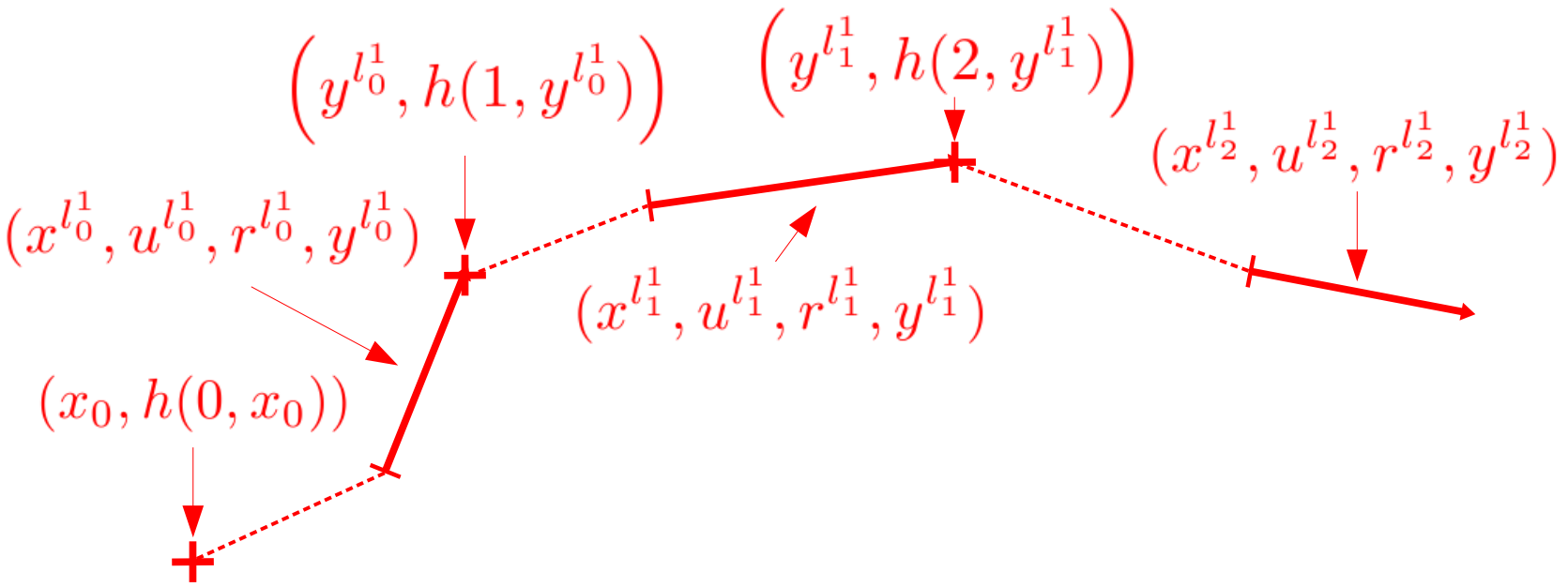
$$\mathcal{G} = \mathcal{G} \setminus \{(x^{l_1^1}, u^{l_1^1}, r^{l_1^1}, y^{l_1^1})\}$$

The MFMC algorithm



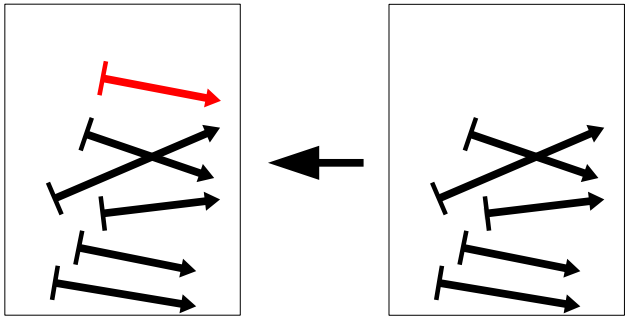
$$\mathcal{G} = \mathcal{G} \setminus \{(x^{l_1^1}, u^{l_1^1}, r^{l_1^1}, y^{l_1^1})\}$$

The MFMC algorithm

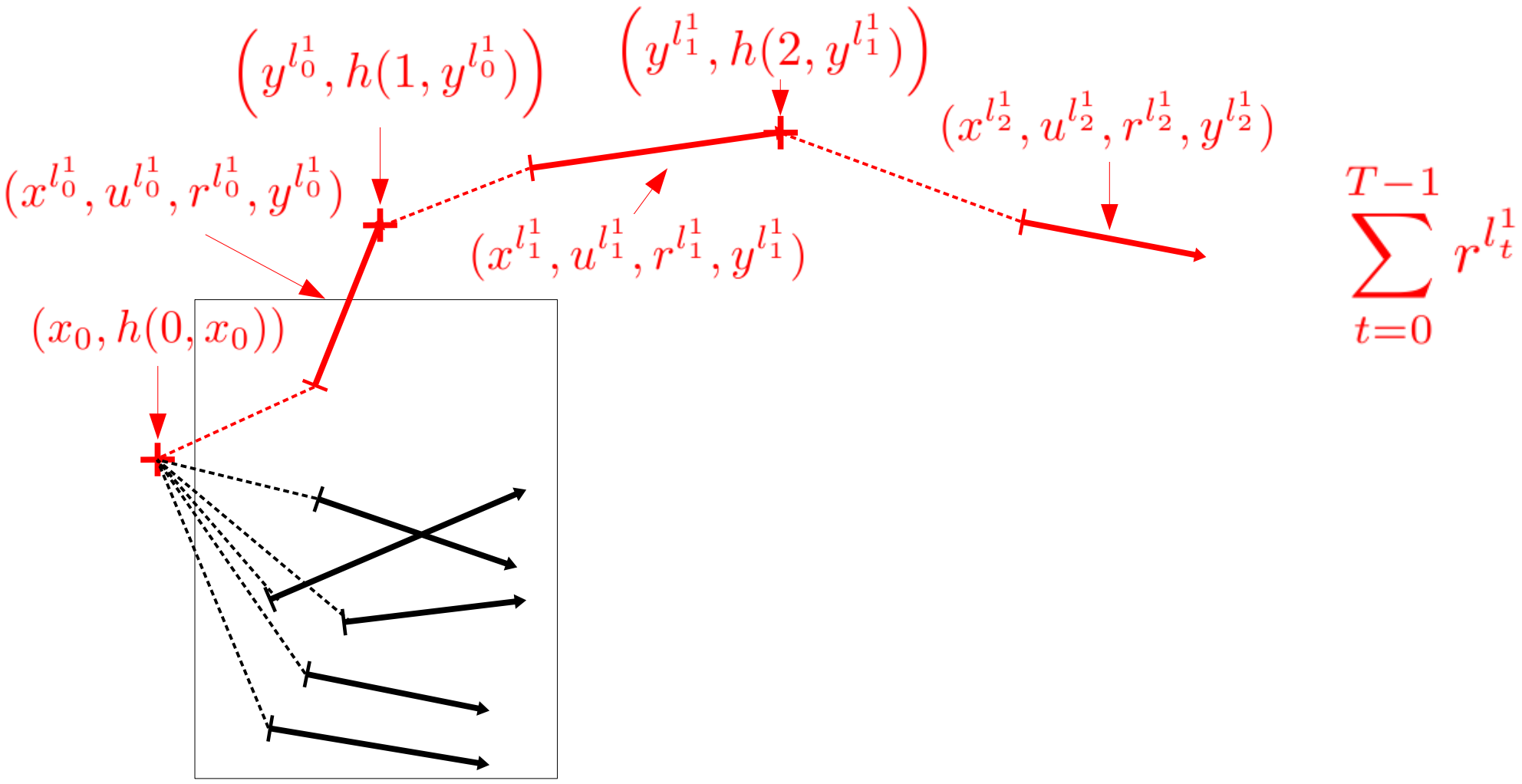


$$\sum_{t=0}^{T-1} r^{l_t^1}$$

$$\mathcal{G} = \mathcal{G} \setminus \{(x^{l_2^1}, u^{l_2^1}, r^{l_2^1}, y^{l_2^1})\}$$

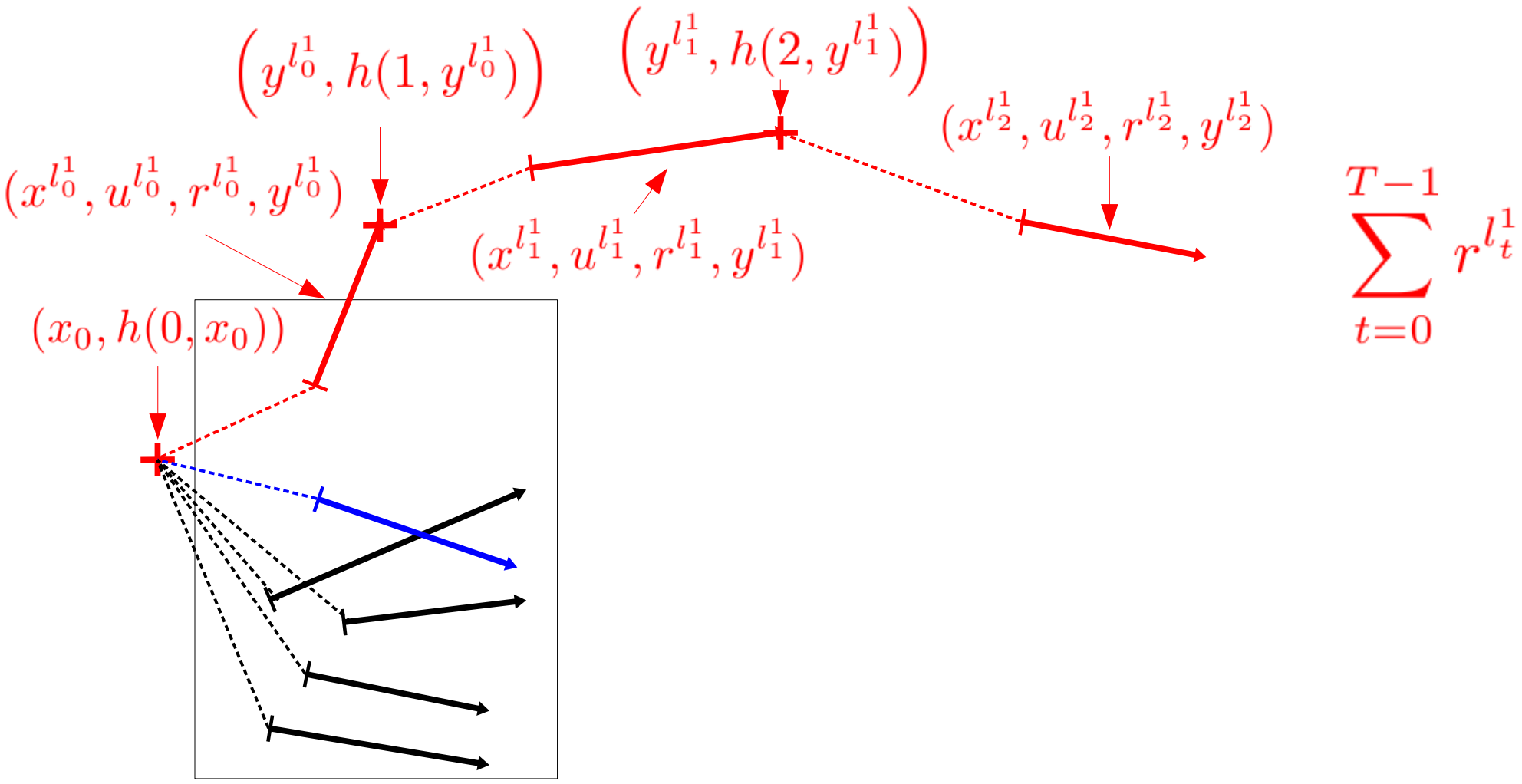


The MFMC algorithm



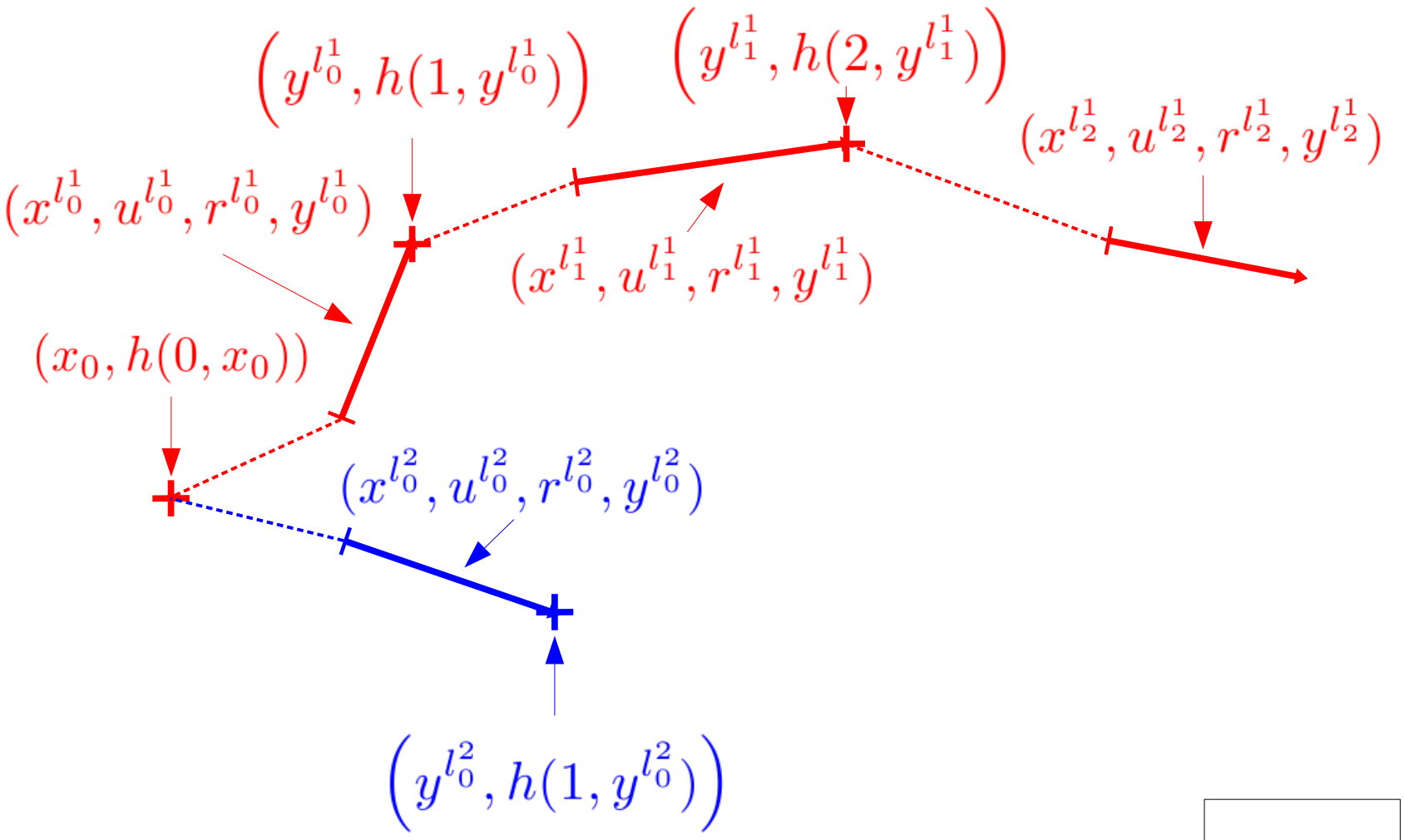
$$\mathcal{G} = \mathcal{G} \setminus \{(x^{l_2^1}, u^{l_2^1}, r^{l_2^1}, y^{l_2^1})\}$$

The MFMC algorithm



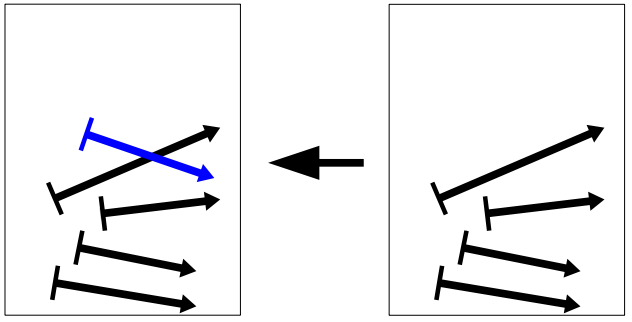
$$\mathcal{G} = \mathcal{G} \setminus \{(x^{l_2^1}, u^{l_2^1}, r^{l_2^1}, y^{l_2^1})\}$$

The MFMC algorithm

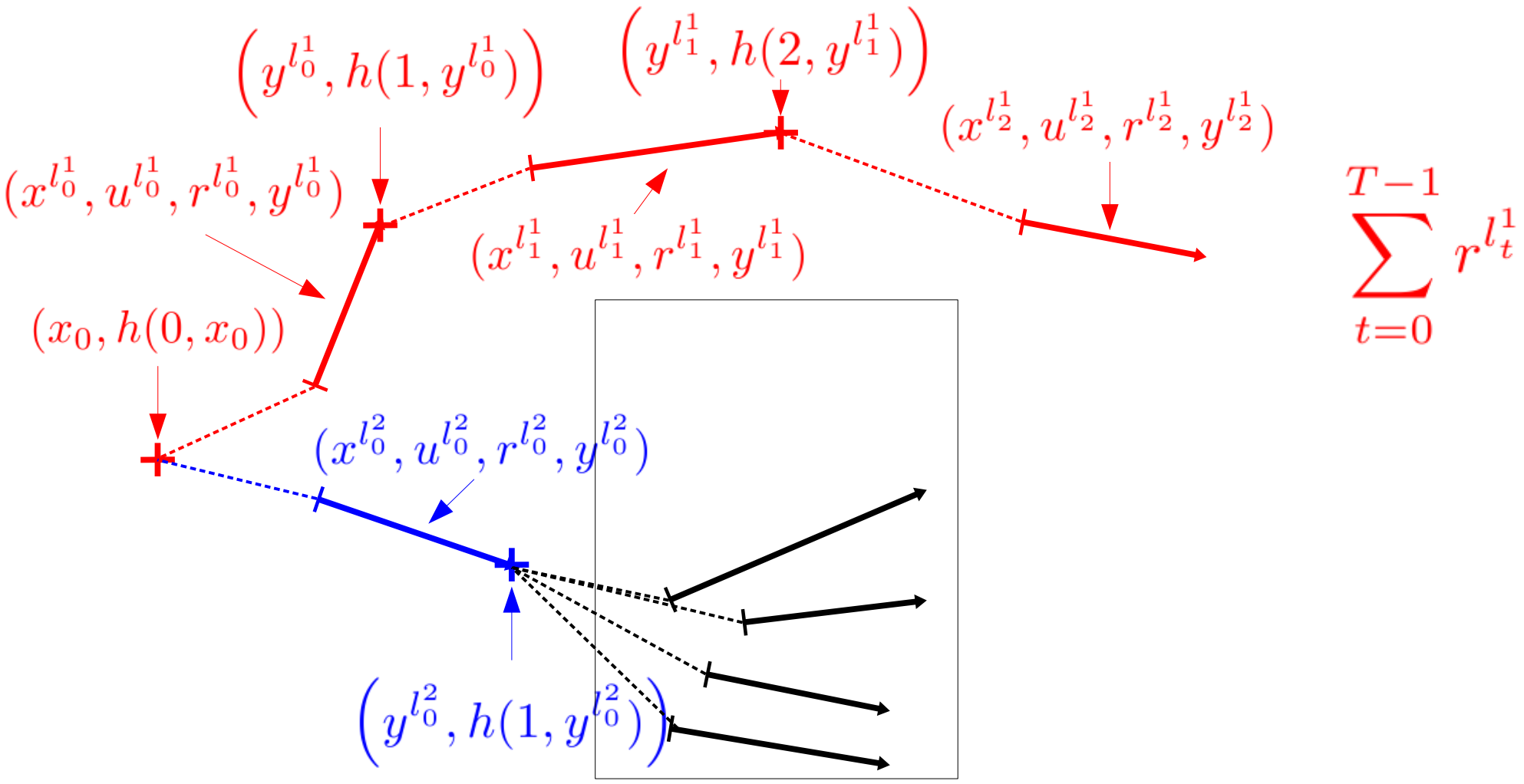


$$\sum_{t=0}^{T-1} r^{l_t^1}$$

$$\mathcal{G} = \mathcal{G} \setminus \{(x^{l_0^2}, u^{l_0^2}, r^{l_0^2}, y^{l_0^2})\}$$

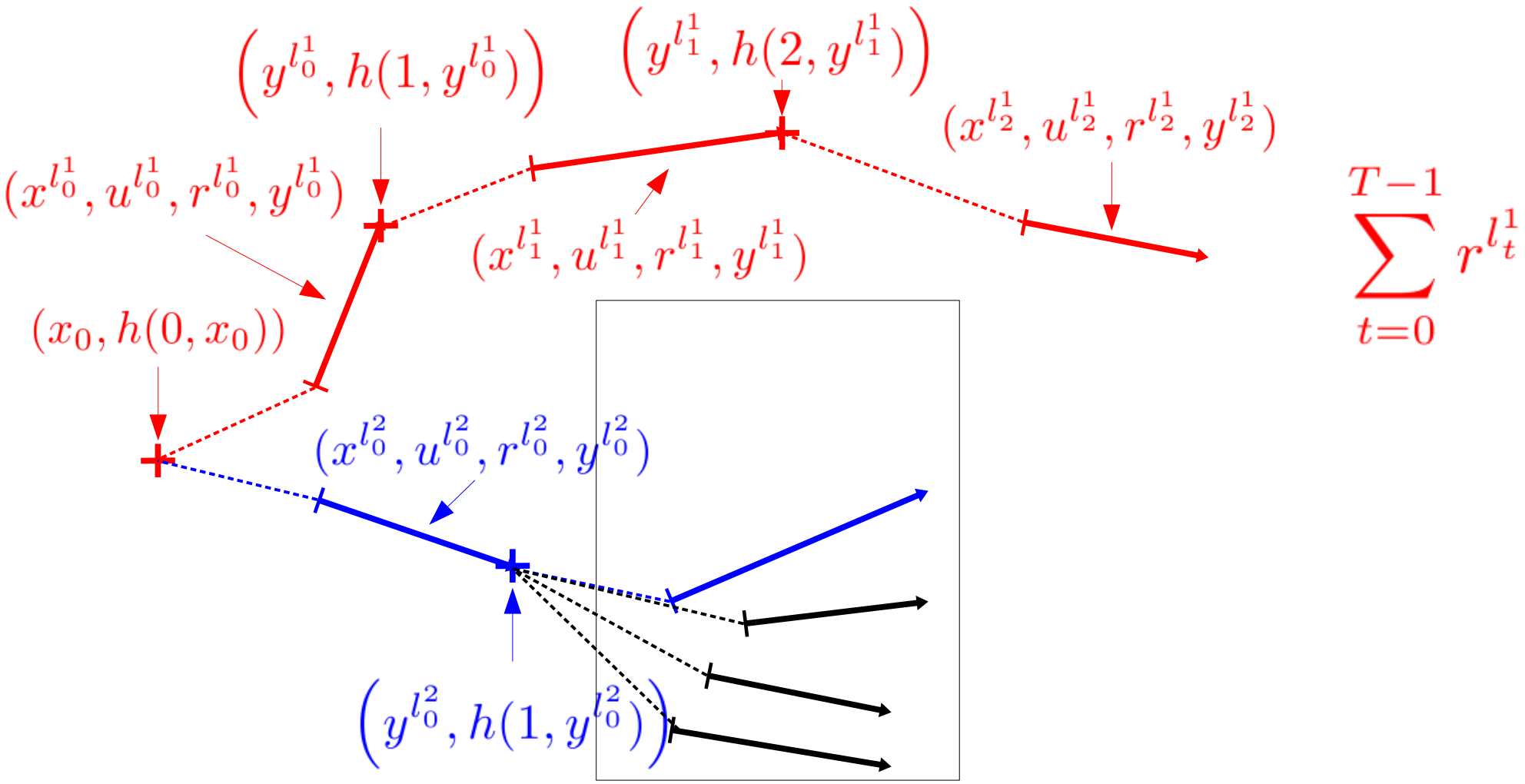


The MFMC algorithm



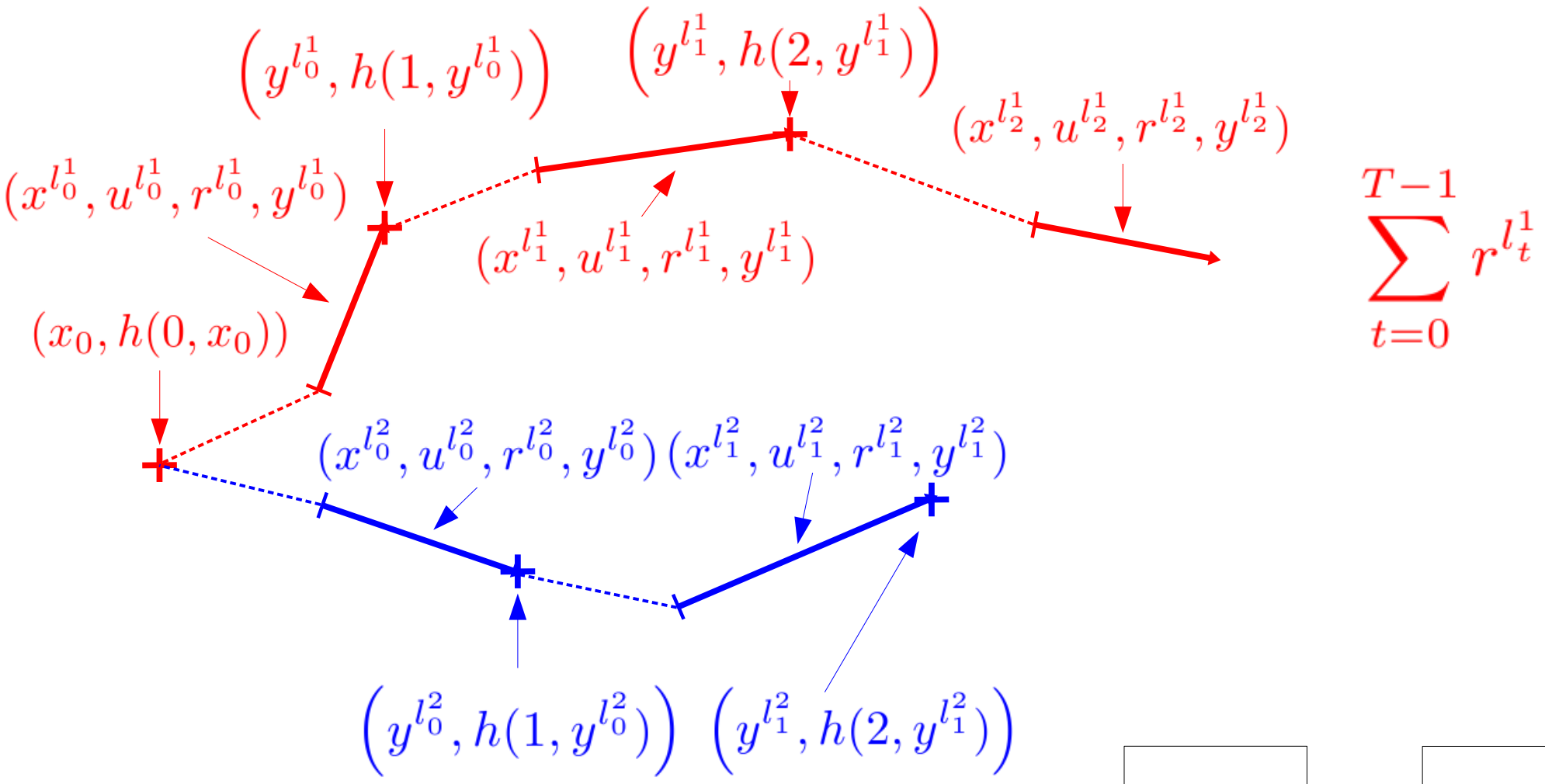
$$\mathcal{G} = \mathcal{G} \setminus \{(x^{l_0^2}, u^{l_0^2}, r^{l_0^2}, y^{l_0^2})\}$$

The MFMC algorithm

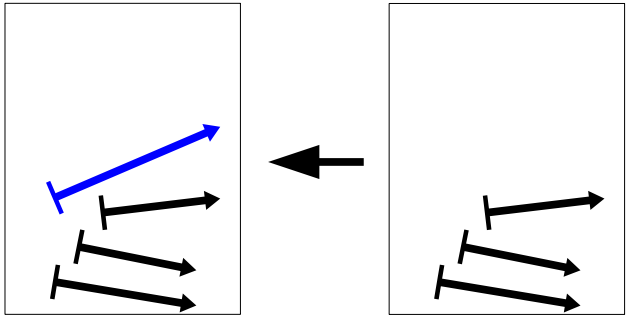


$$\mathcal{G} = \mathcal{G} \setminus \{(x^{l_0^2}, u^{l_0^2}, r^{l_0^2}, y^{l_0^2})\}$$

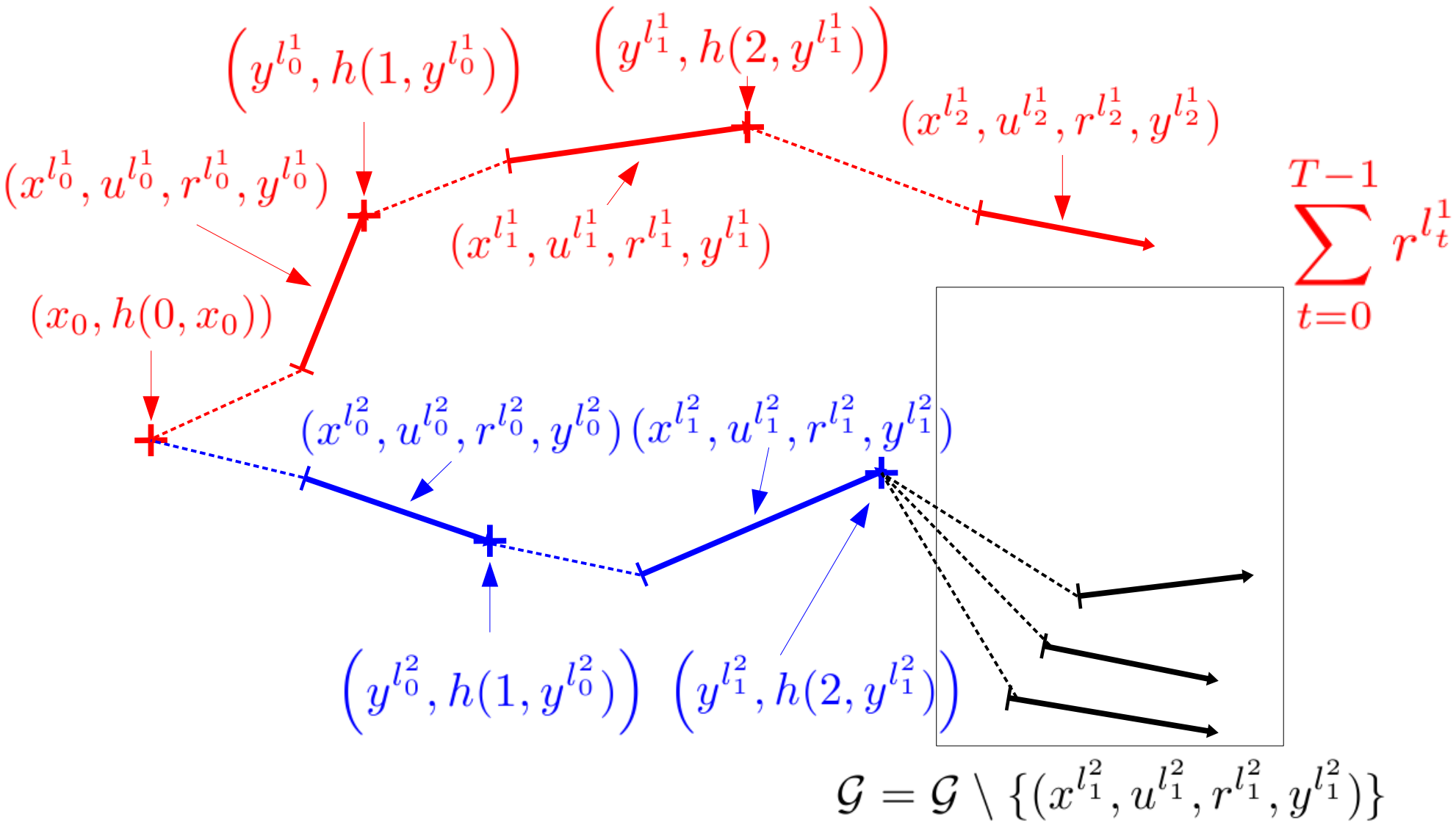
The MFMC algorithm



$$\mathcal{G} = \mathcal{G} \setminus \{(x^{l_1^2}, u^{l_1^2}, r^{l_1^2}, y^{l_1^2})\}$$

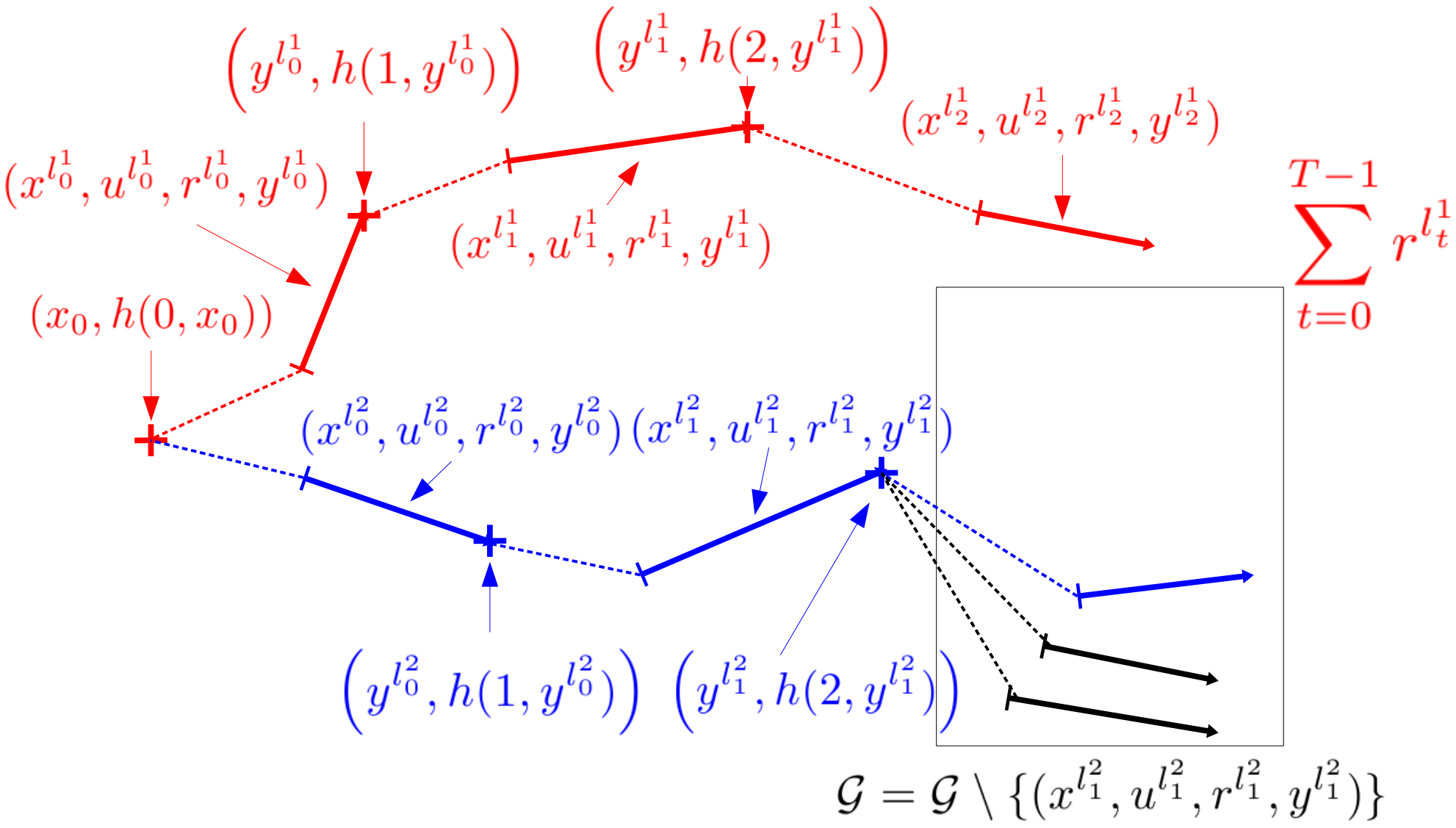


The MFMC algorithm



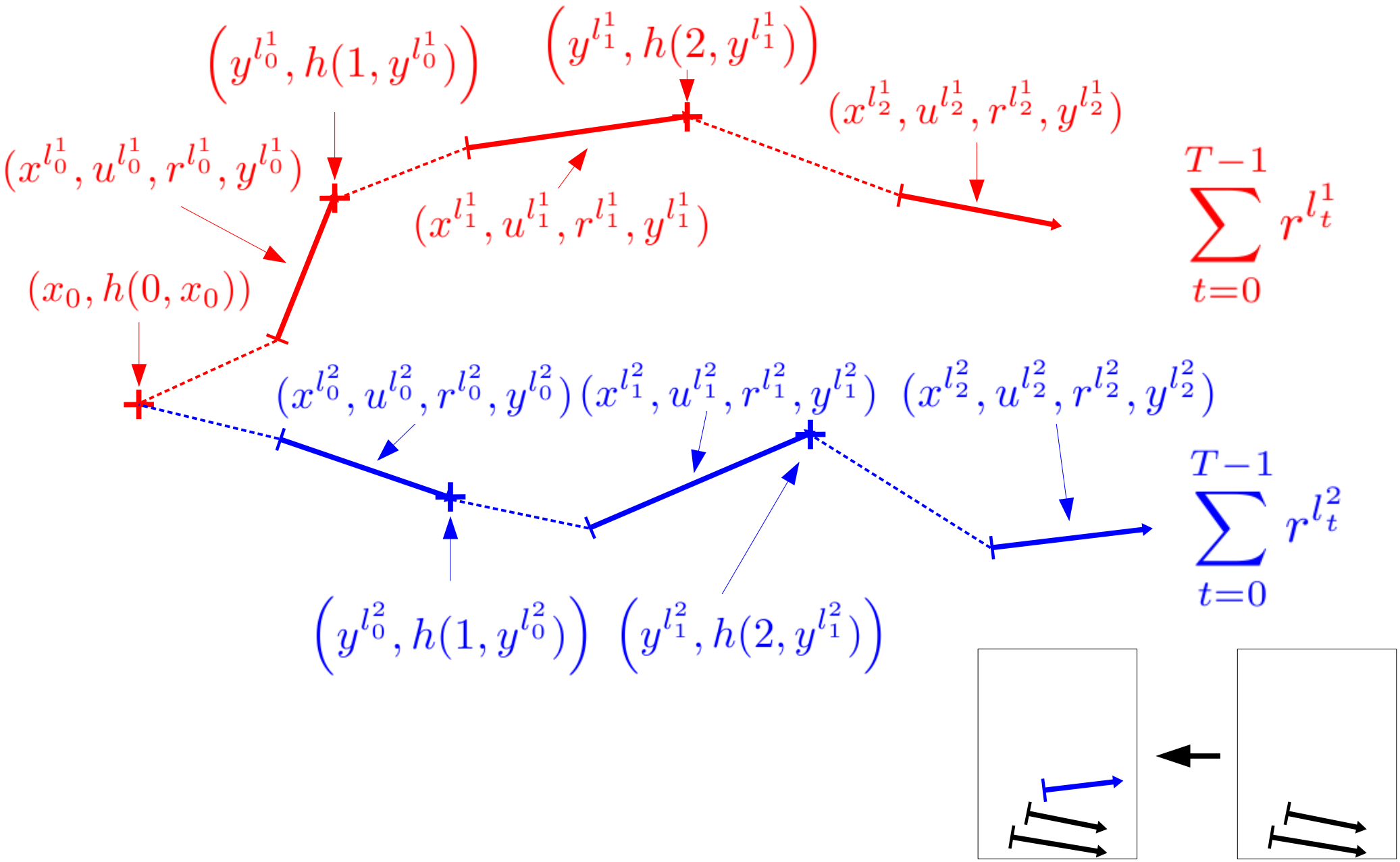
$$\mathcal{G} = \mathcal{G} \setminus \{(x^{l_1^2}, u^{l_1^2}, r^{l_1^2}, y^{l_1^2})\}$$

The MFMC algorithm

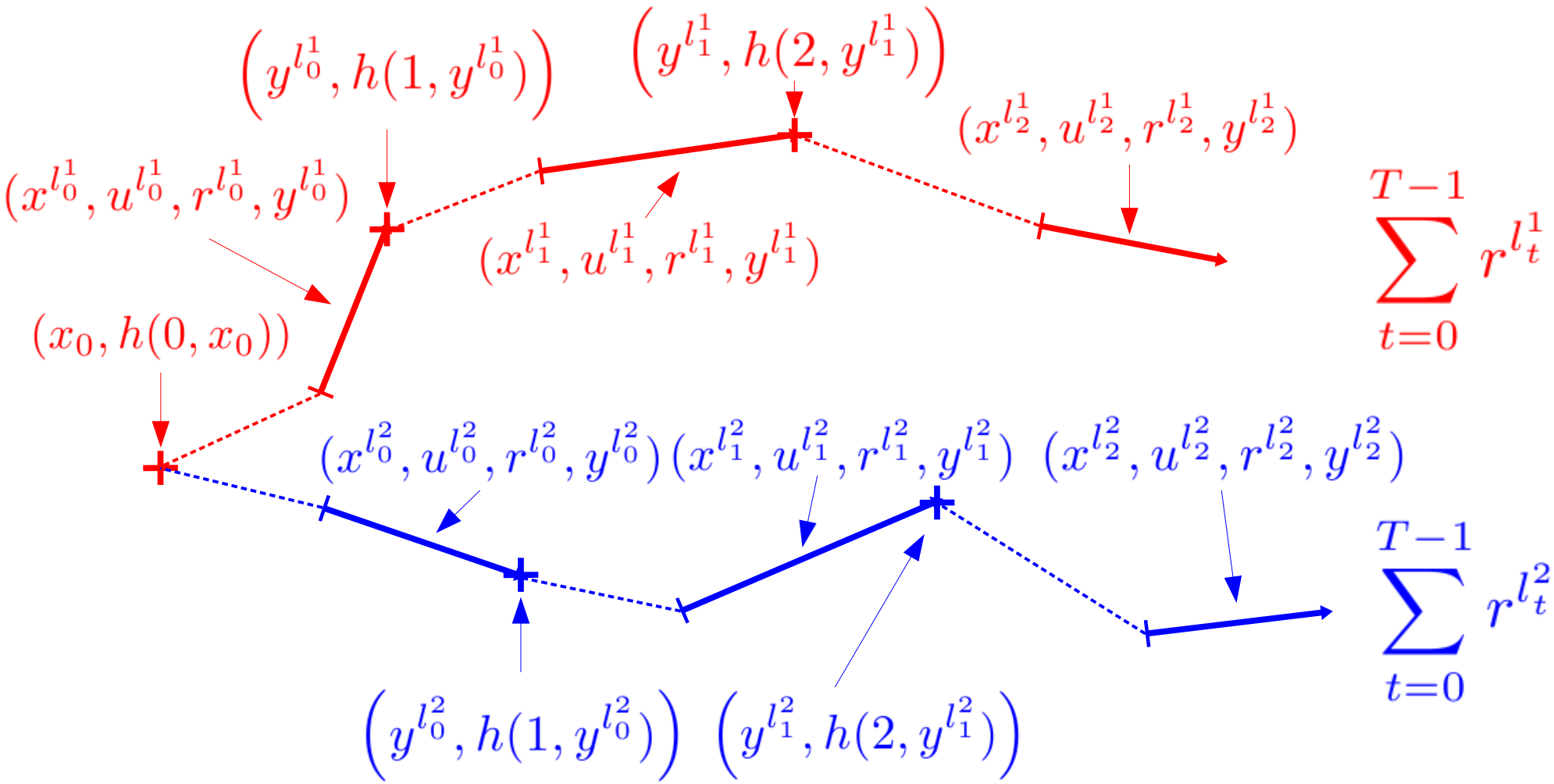


$$\mathcal{G} = \mathcal{G} \setminus \{(x^{l_1^2}, u^{l_1^2}, r^{l_1^2}, y^{l_1^2})\}$$

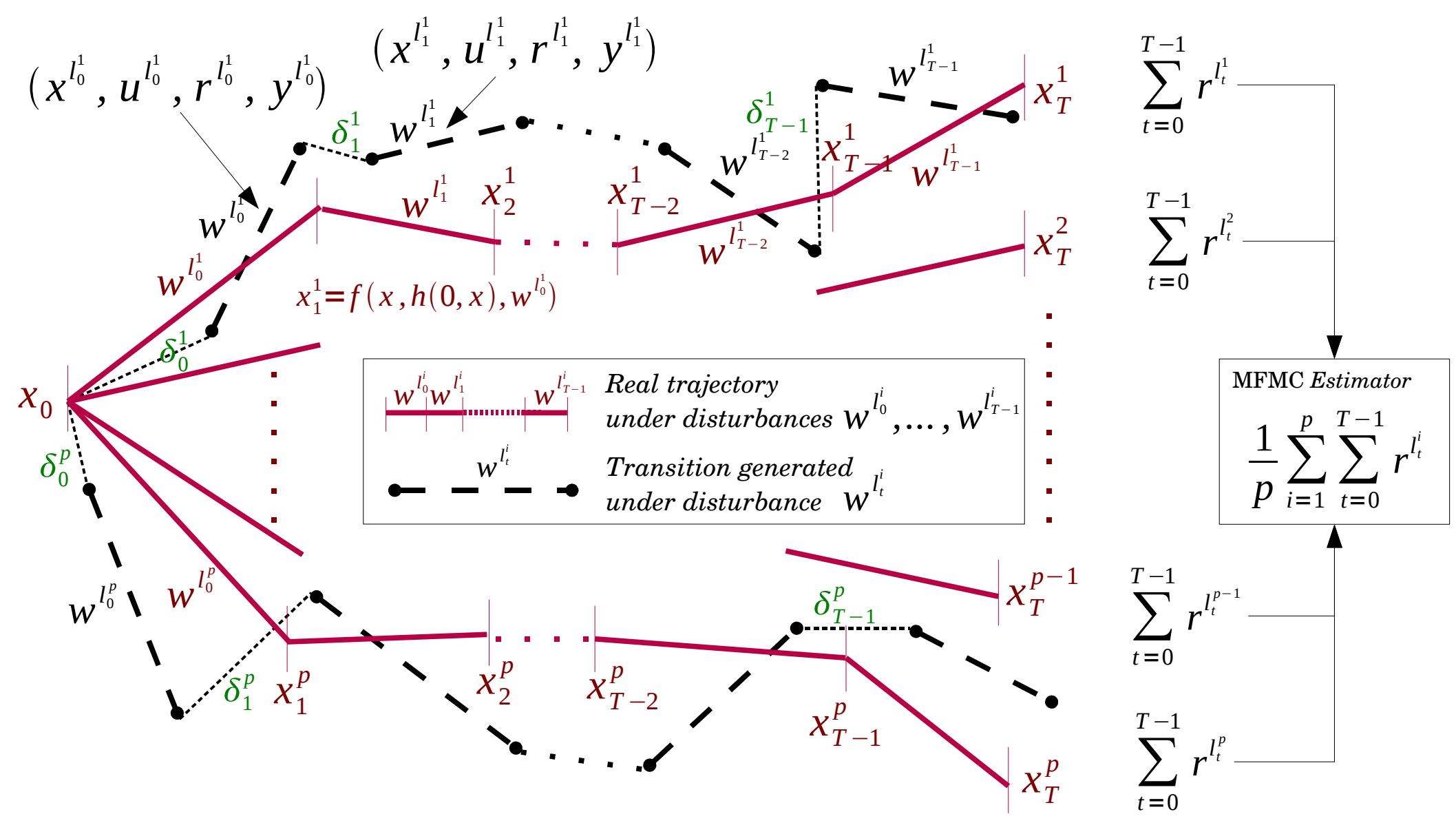
The MFMC algorithm



The MFMC algorithm



$$\mathfrak{M}_2^h(\mathcal{F}_8, x_0) = \frac{1}{2}\left(\sum_{t=0}^{2} r^{l_t^1} + \sum_{t=0}^{2} r^{l_t^2}\right)$$


$(x^{l_0^1}, u^{l_0^1}, r^{l_0^1}, y^{l_0^1})$
$(x^{l_1^1}, u^{l_1^1}, r^{l_1^1}, y^{l_1^1})$
x_0
$w^{l_0^1}$
δ_1^1
$w^{l_1^1}$
x_1^1
$w^{l_{T-2}^1}$
δ_{T-1}^1
$w^{l_{T-1}^1}$
x_{T-2}^1
x_{T-1}^1
x_T^1
x_2^1
$x_1^1 = f(x, h(0,x), w^{l_0^1})$
δ_0^1
x_T^2
δ_0^p
$w^{l_0^p}$
δ_1^p
x_1^p
x_2^p
x_{T-2}^p
δ_{T-1}^p
x_{T-1}^p
x_T^{p-1}
x_T^p
$w^{l_0^i} w^{l_1^i} \dots w^{l_{T-1}^i}$
Real trajectory under disturbances $w^{l_0^i}, \dots, w^{l_{T-1}^i}$
$w^{l_t^i}$
Transition generated under disturbance $w^{l_t^i}$
$\sum_{t=0}^{T-1} r^{l_t^1}$
$\sum_{t=0}^{T-1} r^{l_t^2}$
$\sum_{t=0}^{T-1} r^{l_t^{p-1}}$
$\sum_{t=0}^{T-1} r^{l_t^p}$
MFMC *Estimator*
$\frac{1}{p} \sum_{i=1}^{p} \sum_{t=0}^{T-1} r^{l_t^i}$

What are we going to analyse?

1. we first abstract away from the given sample $\mathcal{F}_n$ by instead considering an ensemble of samples of pairs which are "compatible" with $\mathcal{F}_n$ in the following sense: from the sample

$$\mathcal{F}_n = \left\{ \left(x^l, u^l, r^l, y^l \right) \right\}_{l=1}^{n}, \tag{6.16}$$

we keep only the sample

$$\mathcal{P}_n = \left\{ \left(x^l, u^l \right) \right\}_{l=1}^{n} \in (\mathcal{X} \times \mathcal{U})^n \tag{6.17}$$

of state-action pairs, and we then consider the ensemble of samples of one-step transitions of size n that could be generated by completing each pair (x^l, u^l) of $\mathcal{P}_n$ by drawing for each l a disturbance signal w^l at random from $p_{\mathcal{W}}(.)$, and by recording the resulting values of $f(x^l, u^l, w^l)$ and $\rho(x^l, u^l, w^l)$. We denote by $\tilde{\mathcal{F}}_n$ one such "random" set of one-step transitions defined by a random draw of n disturbance signals $w^l \quad l = 1 \ldots n$. The sample of one-step transitions $\mathcal{F}_n$ is thus a realization of the random set $\tilde{\mathcal{F}}_n$;

What are we going to analyse?

2. we then study the distribution of our estimator $\mathfrak{M}_p^h(\tilde{\mathcal{F}}_n, x_0)$, seen as a function of the random set $\tilde{\mathcal{F}}_n$; in order to characterize this distribution, we express its bias and its variance as a function of a measure of the density of the sample $\mathcal{P}_n$, defined by its "$k-$sparsity"; this is the smallest radius such that all Δ-balls in $\mathcal{X} \times \mathcal{U}$ of this radius contain at least k elements from $\mathcal{P}_n$. The use of this notion implies that the space $\mathcal{X} \times \mathcal{U}$ is bounded (when measured using the distance metric Δ).

Definitions and assumptions

Assumption 6.4.5 (Lipschitz continuity of the functions f, ρ and h)
We assume that the dynamics f, the reward function ρ and the policy h are Lipschitz continuous, i.e.,
$\exists L_f, L_\rho, L_h \in \mathbb{R}^+ : \forall\, (x, x', u, u', w) \in \mathcal{X}^2 \times \mathcal{U}^2 \times \mathcal{W}, \forall t \in \{0, \ldots, T-1\},$

$$\begin{aligned} \|f(x,u,w) - f(x',u',w)\|_{\mathcal{X}} &\leq L_f(\|x - x'\|_{\mathcal{X}} + \|u - u'\|_{\mathcal{U}}), && (6.18)\\ |\rho(x,u,w) - \rho(x',u',w)| &\leq L_\rho(\|x - x'\|_{\mathcal{X}} + \|u - u'\|_{\mathcal{U}}), && (6.19)\\ \|h(t,x) - h(t,x')\|_{\mathcal{U}} &\leq L_h\|x - x'\|_{\mathcal{X}}\ , && (6.20) \end{aligned}$$

where $\|.\|_{\mathcal{X}}$ and $\|.\|_{\mathcal{U}}$ denote the chosen norms over the spaces $\mathcal{X}$ and $\mathcal{U}$, respectively.

Definitions and assumptions

Definition 6.4.6 (Distance metric Δ)
$\forall (x, x', u, u') \in \mathcal{X}^2 \times \mathcal{U}^2$,

$$\Delta((x,u),(x',u')) = (\|x - x'\|_{\mathcal{X}} + \|u - u'\|_{\mathcal{U}}) \, . \tag{6.21}$$

Definition 6.4.7 ($k-$sparsity of a sample $\mathcal{P}_n$)
We suppose that $\mathcal{X} \times \mathcal{U}$ is bounded when measured using the distance metric Δ, and, given $k \in \mathbb{N}_0$ with $k \leq n$, we define the $k-$sparsity, $\alpha_k(\mathcal{P}_n)$ of the sample $\mathcal{P}_n$ by

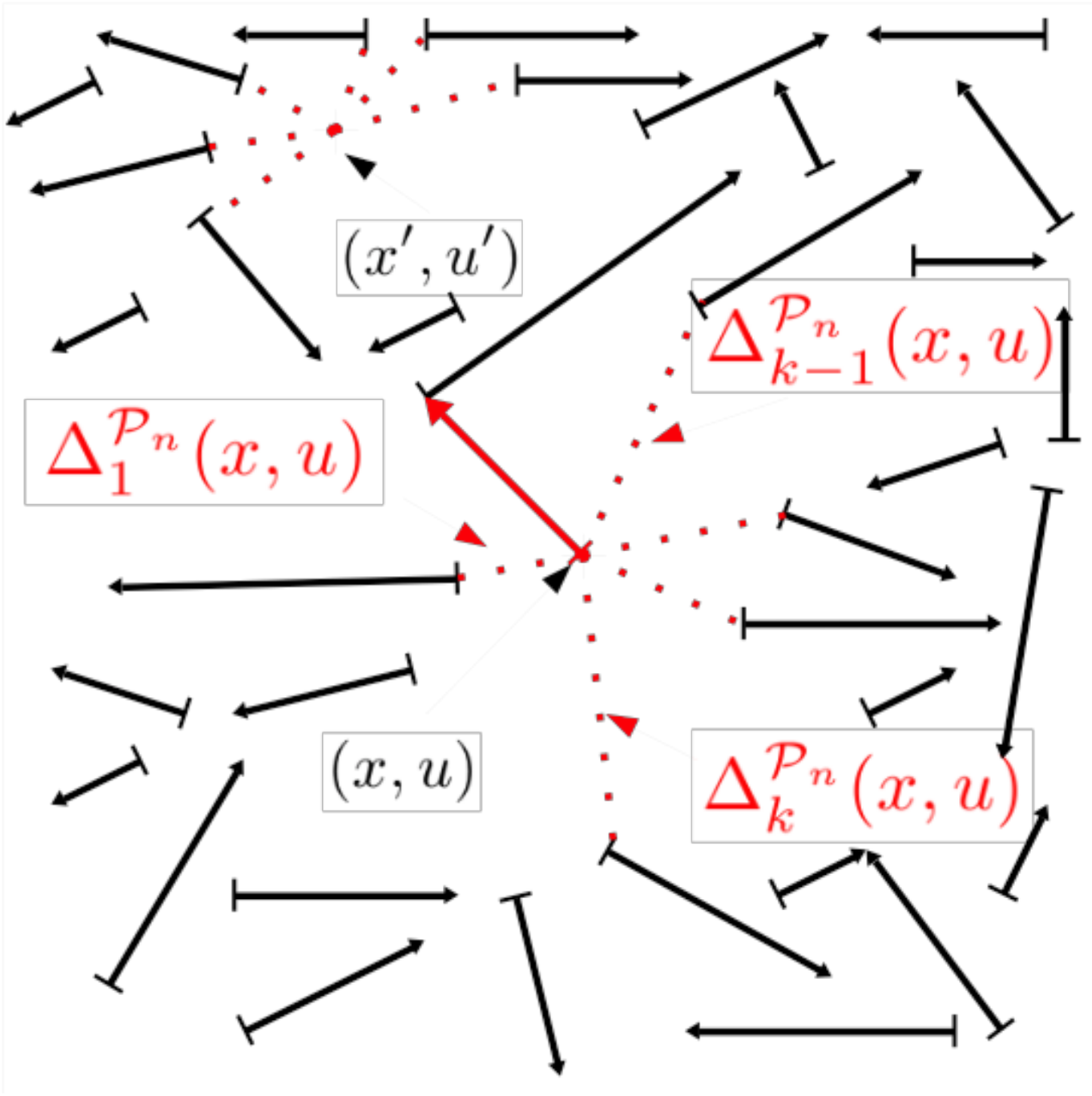
$$\alpha_k(\mathcal{P}_n) = \sup_{(x,u) \in \mathcal{X} \times \mathcal{U}} \left\{ \Delta_k^{\mathcal{P}_n}(x,u) \right\} , \tag{6.22}$$

where $\Delta_k^{\mathcal{P}_n}(x,u)$ denotes the distance of (x,u) to its $k-$th nearest neighbor (using the distance metric Δ) in the $\mathcal{P}_n$ sample.

Definitions and assumptions

The k-dispersion can be seen as the smallest radius such that all Δ-balls in X×U contain at least k elements from

$$\mathcal{P}_n = [(x^l, u^l)]_{l=1}^n$$



Bias

Definition 6.4.8 (Expected value of the MFMC estimator)
$\forall x_0 \in \mathcal{X}$,

$$E^h_{p,\mathcal{P}_n}(x_0) = \mathop{\mathbb{E}}_{w^1,\dots,w^n \sim p_{\mathcal{W}}(.)} \left[\mathfrak{M}^h_p(\tilde{\mathcal{F}}_n, x_0)\right]. \tag{6.23}$$

Theorem 6.4.9 (Bias of the MFMC estimator)

$$\forall x_0 \in \mathcal{X}, \qquad \left|J^h(x_0) - E^h_{p,\mathcal{P}_n}(x_0)\right| \leq C\alpha_{pT}(\mathcal{P}_n) \tag{6.24}$$

$$\textit{with } C = L_\rho \sum_{t=0}^{T-1} \sum_{i=0}^{T-t-1} \left[L_f(1+L_h)\right]^i . \tag{6.25}$$

Bias - proof

Vector of disturbances

$$\Omega = [\Omega(0), \ldots, \Omega(T-1)] \in \mathcal{W}^T, \tag{6.26}$$

Definition 6.4.10 (Ω-disturbed state-action value function)
$\forall t \in \{0, \ldots, T-1\}, \forall (x, u) \in \mathcal{X} \times \mathcal{U}, \forall \Omega \in \mathcal{W}^T,$

$$Q_{T-t}^{h,\Omega}(x, u) = \rho(x, u, \Omega(t)) + \sum_{t'=t+1}^{T-1} \rho(x_{t'}, h(t', x_{t'}), \Omega(t')) \tag{6.27}$$

with

$$x_{t+1} = f(x, u, \Omega(t)) \tag{6.28}$$

and

$$\forall t' \in \{t+1, \ldots, T-1\}, x_{t'+1} = f(x_{t'}, h(t', x_{t'}), \Omega(t')). \tag{6.29}$$

Bias - proof

Definition 6.4.11 (Expected return given Ω)
$\forall x_0 \in \mathcal{X}, \forall \Omega \in \mathcal{W}^T$,

$$\mathbb{E}[R^h(x_0)|\Omega] = \mathop{\mathbb{E}}_{w_0,\dots,w_{T-1}\sim p_{\mathcal{W}}(.)}[R^h(x_0)|w_0 = \Omega(0), \dots, w_{T-1} = \Omega(T-1)]. \tag{6.30}$$

Proposition 6.4.12
$\forall x_0 \in \mathcal{X}, \forall \Omega \in \mathcal{W}^T$,

$$\mathbb{E}[R^h(x_0)|\Omega] = Q_T^{h,\Omega}(x_0, h(0, x_0)) \ . \tag{6.31}$$

Proposition 6.4.13
$\forall (x, u) \in \mathcal{X} \times \mathcal{U}, \forall \Omega \in \mathcal{W}^T$,

$$\begin{aligned} Q_{T-t+1}^{h,\Omega}(x,u) &= \rho(x, u, \Omega(t-1)) \\ &+ Q_{T-t}^{h,\Omega}\big(f\big(x, u, \Omega(t-1)\big), h\big(t, f(x, u, \Omega(t-1))\big)\big) \ . \end{aligned} \tag{6.32}$$

Bias - proof

Lemma 6.4.14 (Lipschitz Continuity of $Q_{T-t}^{h,\Omega}$)
$\forall t \in \{0, \dots, T-1\}, \forall (x, x', u, u') \in \mathcal{X}^2 \times \mathcal{U}^2,$

$$\left| Q_{T-t}^{h,\Omega}(x,u) - Q_{T-t}^{h,\Omega}(x',u') \right| \leq L_{Q_{T-t}} \Delta((x,u),(x',u')) \tag{6.33}$$

with

$$L_{Q_{T-t}} = L_\rho \sum_{i=0}^{T-t-1} \left[L_f (1 + L_h) \right]^i. \tag{6.34}$$

Proof: by induction.

Bias - proof

Proof. We denote by $\mathcal{H}(T-t)$ the proposition:
$\mathcal{H}(T-t) : \forall (x, x', u, u') \in \mathcal{X}^2 \times \mathcal{U}^2,$

$$\left| Q_{T-t}^{h,\Omega}(x,u) - Q_{T-t}^{h,\Omega}(x',u') \right| \leq L_{Q_{T-t}} \Delta((x,u),(x',u')) \ . \tag{6.35}$$

We prove by induction that $\mathcal{H}(T-t)$ is true $\forall t \in \{0, \dots, T-1\}$. For the sake of conciseness, we denote use the notation

$$\Delta_{T-t}^{Q} = \left| Q_{T-t}^{h,\Omega}(x,u) - Q_{T-t}^{h,\Omega}(x',u') \right| \ . \tag{6.36}$$

- **Basis:** $t = T-1$

We have

$$\Delta_1^Q = |\rho(x,u,\Omega(T-1)) - \rho(x',u',\Omega(T-1)|, \tag{6.37}$$

and the Lipschitz continuity of ρ allows to write

$$\Delta_1^Q \leq L_\rho \big(\|x-x'\|_{\mathcal{X}} + \|u-u'\|_{\mathcal{U}} \big) \ = L_\rho \Delta((x,u),(x',u')) \ . \tag{6.38}$$

This proves $\mathcal{H}(1)$.

Proof - bias

- **Induction step:** We suppose that $\mathcal{H}(T-t)$ is true, $1 \leq t \leq T-1$.

Using Equation 6.4.13, one has

$$\begin{aligned}
\Delta^{Q}_{T-t+1} &= \left|Q^{h,\Omega}_{T-t+1}(x,u) - Q^{h,\Omega}_{T-t+1}(x',u')\right| && (6.39)\\
&= \Big|\rho(x,u,\Omega(t-1)) - \rho(x',u',\Omega(t-1)) \\
&+ Q^{h,\Omega}_{T-t}(f(x,u,\Omega(t-1)), h(t,f(x,u,\Omega(t-1)))) \\
&- Q^{h,\Omega}_{T-t}(f(x',u',\Omega(t-1)), h(t,f(x',u',\Omega(t-1))))\Big| && (6.40)
\end{aligned}$$

and, from there,

$$\begin{aligned}
\Delta^{Q}_{T-t+1} &\leq \left|\rho(x,u,\Omega(t-1)) - \rho(x',u',\Omega(t-1))\right| \\
&+ \left|Q^{h,\Omega}_{T-t}(f(x,u,\Omega(t-1)), h(t,f(x,u,\Omega(t-1))))\right| \\
&- Q^{h,\Omega}_{T-t}(f(x',u',\Omega(t-1)), h(t,f(x',u',\Omega(t-1))))\Big|. && (6.41)
\end{aligned}$$

Proof - bias

$\mathcal{H}(T-t)$ and the Lipschitz continuity of ρ give

$$\begin{aligned}\Delta^{Q}_{T-t+1} &\leq L_{\rho}\Delta((x,u),(x',u')) \\ &+ L_{Q_{T-t}}\Delta((f(x,u,\Omega(t-1)),h(t,f(x,u,\Omega(t-1)))), \\ &\quad (f(x',u',\Omega(t-1)),h(t,f(x',u',\Omega(t-1))))) \,. \end{aligned} \tag{6.42}$$

Using the Lipschitz continuity of f and h, we have

$$\begin{aligned}\Delta^{Q}_{T-t+1} &\leq L_{\rho}\Delta((x,u),(x',u')) \\ &+ L_{Q_{T-t}}\big(L_{f}\Delta((x,u),(x',u')) + L_{h}L_{f}\Delta((x,u),(x',u'))\big), \end{aligned} \tag{6.43}$$

and, from there,

$$\Delta^{Q}_{T-t+1} \leq L_{Q_{T-t+1}}\Delta((x,u),(x',u')) \tag{6.44}$$

since

$$L_{Q_{T-t+1}} \doteq L_{\rho} + L_{Q_{T-t}}L_{f}(1+L_{h}). \tag{6.45}$$

This proves $\mathcal{H}\,(T-t+1)$ and ends the proof. ■

Proof - bias

Definition 6.4.15 (Disturbance vector associated with a broken trajectory)
Given a broken trajectory

$$\tau^i = \left[\left(x^{l_t^i}, u^{l_t^i}, r^{l_t^i}, y^{l_t^i}\right)\right]_{t=0}^{T-1} \tag{6.46}$$

we denote by Ω^{τ^i} *its associated disturbance vector*

$$\Omega^{\tau^i} = \left[w^{l_0^i}, \ldots, w^{l_{T-1}^i}\right], \tag{6.47}$$

i.e. the vector made of the T *unknown disturbances that affected the generation of the one-step transitions* $\left(x^{l_t^i}, u^{l_t^i}, r^{l_t^i}, y^{l_t^i}\right)$ *(cf. first item of Section 6.4.3).*

Proof - bias

Lemma 6.4.16 (Bounds on the expected return given Ω)
$\forall x_0 \in \mathcal{X}, \forall i \in \{1, \dots, p\}$,

$$b^h(\tau^i, x_0) \leq \mathbb{E}\left[R^h(x_0)|\Omega^{\tau^i}\right] \leq a^h(\tau^i, x_0) \ , \tag{6.48}$$

with

$$b^h(\tau^i, x_0) = \sum_{t=0}^{T-1} \left[r^{l_t^i} - L_{Q_{T-t}} \delta_t^i\right] \ , \tag{6.49}$$

$$a^h(\tau^i, x_0) = \sum_{t=0}^{T-1} \left[r^{l_t^i} + L_{Q_{T-t}} \delta_t^i\right] \ , \tag{6.50}$$

$$\delta_t^i = \Delta\left(\left(x^{l_t^i}, u^{l_t^i}\right), \left(y^{l_{t-1}^i}, h\left(t, y^{l_{t-1}^i}\right)\right)\right) \ , \forall t \in \{0, \dots, T-1\} \ , \tag{6.51}$$

$$y^{l_{-1}^i} = x_0, \forall i \in \{1, \dots, p\}. \tag{6.52}$$

Proof - bias

Proof. Let us first prove the lower bound. With $u_0 = h(0, x_0)$, the Lipschitz continuity of $Q_T^{h,\Omega^{\tau^i}}$ gives

$$\left|Q_T^{h,\Omega^{\tau^i}}(x_0, u_0) - Q_T^{h,\Omega^{\tau^i}}(x^{l_0^i}, u^{l_0^i})\right| \leq L_{Q_T} \Delta\left((x_0, u_0), \left(x^{l_0^i}, u^{l_0^i}\right)\right) . \tag{6.53}$$

According to Proposition (6.4.12),

$$Q_T^{h,\Omega^{\tau^i}}(x_0, u_0) = \mathbb{E}\left[R^h(x_0)|\Omega^{\tau^i}\right] . \tag{6.54}$$

Thus,

$$\begin{aligned} &\left|\mathbb{E}\left[R^h(x_0)|\Omega^{\tau^i}\right] - Q_T^{h,\Omega^{\tau^i}}\left(x^{l_0^i}, u^{l_0^i}\right)\right| \\ &= \left|Q_T^{h,\Omega^{\tau^i}}(x_0, h(0, x_0)) - Q_T^{h,\Omega^{\tau^i}}\left(x^{l_0^i}, u^{l_0^i}\right)\right| &(6.55)\\ &\leq L_{Q_T} \Delta\left((x_0, h(0, x_0)), \left(x^{l_0^i}, u^{l_0^i}\right)\right) . &(6.56)\end{aligned}$$

It follows that

$$Q_T^{h,\Omega^{\tau^i}}\left(x^{l_0^i}, u^{l_0^i}\right) - L_{Q_T}\delta_0^i \leq \mathbb{E}\left[R^h(x_0)|\Omega^{\tau^i}\right] . \tag{6.57}$$

Proof - bias

Using Equation (6.4.13) we have

$$\begin{aligned} Q_T^{h,\Omega^{\tau^i}}\left(x^{l_0^i}, u^{l_0^i}\right) &= \rho\left(x^{l_0^i}, u^{l_0^i}, w^{l_0^i}\right) \\ &+ Q_{T-1}^{h,\Omega^{\tau^i}}\left(f\left(x^{l_0^i}, u^{l_0^i}, w^{l_0^i}\right), h\left(1, f\left(x^{l_0^i}, u^{l_0^i}, w^{l_0^i}\right)\right)\right) . \end{aligned} \tag{6.58}$$

By definition of Ω^{τ^i}, we have

$$\rho\left(x^{l_0^i}, u^{l_0^i}, w^{l_0^i}\right) = r^{l_0^i} \tag{6.59}$$

and

$$f\left(x^{l_0^i}, u^{l_0^i}, w^{l_0^i}\right) = y^{l_0^i} . \tag{6.60}$$

From there

$$Q_T^{h,\Omega^{\tau^i}}\left(x^{l_0^i}, u^{l_0^i}\right) = r^{l_0^i} + Q_{T-1}^{h,\Omega^{\tau^i}}\left(y^{l_0^i}, h\left(1, y^{l_0^i}\right)\right) , \tag{6.61}$$

and

$$Q_{T-1}^{h,\Omega^{\tau^i}}\left(y^{l_0^i}, h\left(1, y^{l_0^i}\right)\right) + r^{l_0^i} - L_{Q_T}\delta_0^i \leq \mathbb{E}\left[R^h(x_0)|\Omega^{\tau^i}\right] . \tag{6.62}$$

Proof - bias

The Lipschitz continuity of $Q_{T-1}^{h,\Omega^{\tau^i}}$ gives

$$\begin{aligned}
&\left|Q_{T-1}^{h,\Omega^{\tau^i}}\left(y^{l_0^i}, h\left(1, y^{l_0^i}\right)\right) - Q_{T-1}^{h,\Omega^{\tau^i}}\left(x^{l_1^i}, u^{l_1^i}\right)\right| \\
&\leq L_{Q_{T-1}} \Delta\left(\left(y^{l_0^i}, h\left(1, y^{l_0^i}\right)\right), \left(x^{l_1^i}, u^{l_1^i}\right)\right) \qquad (6.63) \\
&= L_{Q_{T-1}} \delta_1^i, \qquad (6.64)
\end{aligned}$$

which implies that

$$Q_{T-1}^{h,\Omega^{\tau^i}}\left(x^{l_1^i}, u^{l_1^i}\right) - L_{Q_{T-1}}\delta_1^i \leq Q_{T-1}^{h,\Omega^{\tau^i}}\left(y^{l_0^i}, h\left(1, y^{l_0^i}\right)\right) . \qquad (6.65)$$

We therefore have

$$Q_{T-1}^{h,\Omega^{\tau^i}}\left(x^{l_1^i}, u^{l_1^i}\right) + r^{l_0^i} - L_{Q_T}\delta_0^i - L_{Q_{T-1}}\delta_1^i \leq \mathbb{E}\left[R^h(x_0)|\Omega^{\tau^i}\right] . \qquad (6.66)$$

The proof is completed by iterating this derivation. The upper bound is proved similarly. ■

Proof - bias

Lemma 6.4.17
$\forall x_0 \in \mathcal{X}, \forall i \in \{1, \ldots, p\}$,

$$a^h\left(\tau^i, x_0\right) - b^h\left(\tau^i, x_0\right) \leq 2C\alpha_{pT}\left(\mathcal{P}_n\right) \tag{6.67}$$

with

$$C = \sum_{t=0}^{T-1} L_{Q_{T-t}} \ . \tag{6.68}$$

Proof - bias

Proof. By construction of the bounds, one has

$$a^h\left(\tau^i, x_0\right) - b^h\left(\tau^i, x_0\right) = \sum_{t=0}^{T-1} 2L_{Q_{T-t}}\delta_t^i \ . \tag{6.69}$$

The MFMC algorithm chooses $p \times T$ different one-step transitions to build the MFMC estimator by minimizing the distance $\Delta((y^{l_{t-1}^i}, h(t, y^{l_{t-1}^i})), (x^{l_t^i}, u^{l_t^i}))$, so by definition of the k-sparsity of the sample $\mathcal{P}_n$ with $k = pT$, one has

$$\begin{aligned}
\delta_t^i &= \Delta\left(\left(y^{l_{t-1}^i}, h\left(t, y^{l_{t-1}^i}\right)\right), \left(x^{l_t^i}, u^{l_t^i}\right)\right) && (6.70)\\
&\leq \Delta_{pT}^{\mathcal{P}_n}\left(y^{l_{t-1}^i}, h\left(t, y^{l_{t-1}^i}\right)\right) && (6.71)\\
&\leq \alpha_{pT}\left(\mathcal{P}_n\right) \ , && (6.72)
\end{aligned}$$

which ends the proof. ■

Proof - bias

Proof of Theorem 6.4.9 By definition of $a^h(\tau^i, x_0)$ and $b^h(\tau^i, x_0)$, we have

$$\forall i \in \{1, \dots, p\}, \frac{b^h\left(\tau^i, x_0\right) + a^h\left(\tau^i, x_0\right)}{2} = \sum_{t=0}^{T-1} r^{l_t^i} \ . \tag{6.73}$$

Then, according to Lemmas 6.4.16 and 6.4.17, we have $\forall i \in \{1, \dots, p\}$,

$$\begin{aligned}
&\left| \underset{w^1,\dots,w^n \sim p_{\mathcal{W}}(.)}{\mathbb{E}} \left[\mathbb{E}\left[R^h(x_0) | \Omega^{\tau^i} \right] - \sum_{t=0}^{T-1} r^{l_t^i} \right] \right| \\
&\leq \underset{w^1,\dots,w^n \sim p_{\mathcal{W}}(.)}{\mathbb{E}} \left[\left| \mathbb{E}\left[R^h(x_0) | \Omega^{\tau^i} \right] - \sum_{t=0}^{T-1} r^{l_t^i} \right| \right] \qquad (6.74) \\
&\leq C \alpha_{pT}(\mathcal{P}_n) \ . \qquad (6.75)
\end{aligned}$$

Proof - bias

Thus,

$$\left| \frac{1}{p} \sum_{i=1}^{p} \mathop{\mathbb{E}}_{w^1,\ldots,w^n \sim p_{\mathcal{W}}(.)} \left[\mathbb{E}\left[R^h(x_0) | \Omega^{\tau^i} \right] - \sum_{t=0}^{T-1} r^{l_t^i} \right] \right|$$

$$\leq \frac{1}{p} \sum_{i=1}^{p} \left| \mathop{\mathbb{E}}_{w^1,\ldots,w^n \sim p_{\mathcal{W}}(.)} \left[\mathbb{E}\left[R^h(x_0) | \Omega^{\tau^i} \right] - \sum_{t=0}^{T-1} r^{l_t^i} \right] \right| \tag{6.76}$$

$$\leq C\alpha_{pT}\left(\mathcal{P}_n\right) \ , \tag{6.77}$$

which can be reformulated

$$\left| \mathop{\mathbb{E}}_{w^1,\ldots,w^n \sim p_{\mathcal{W}}(.)} \left[\frac{1}{p} \sum_{i=1}^{p} \mathbb{E}\left[R^h(x_0) | \Omega^{\tau^i} \right] \right] - E^h_{p,\mathcal{P}_n}(x_0) \right| \leq C\alpha_{pT}\left(\mathcal{P}_n\right) \ , \tag{6.78}$$

since

$$\frac{1}{p} \sum_{i=1}^{p} \sum_{t=0}^{T-1} r^{l_t^i} = \mathfrak{M}_p^h(\tilde{\mathcal{F}}_n, x_0) \ . \tag{6.79}$$

Proof - bias

Since the MFMC algorithm chooses $p \times T$ different one-step transitions, all the disturbances $\left\{ w^{l_t^i} \right\}_{i=1,t=0}^{i=p,t=T-1}$ are i.i.d. according to $p_{\mathcal{W}}(.)$. For all $i \in \{1, \ldots, p\}$, The law of total expectation gives

$$\begin{aligned}
&\mathop{\mathbb{E}}_{w^{l_0^i},\ldots,w^{l_{T-1}^i} \sim p_{\mathcal{W}}(.)} \left[\mathop{\mathbb{E}}_{w^{l_0^i},\ldots,w^{l_{T-1}^i} \sim p_{\mathcal{W}}(.)} \left[R^h(x_0) | \Omega^{\tau^i} \right] \right] \\
&= \mathop{\mathbb{E}}_{w_0,\ldots,w_{T-1} \sim p_{\mathcal{W}}(.)} \left[R^h(x_0) \right] && (6.80) \\
&= J^h(x_0) \, . && (6.81)
\end{aligned}$$

This ends the proof. ■

Variance

Definition 6.4.18 (Variance of the MFMC estimator)
$\forall x_0 \in \mathcal{X}$,

$$V_{p,\mathcal{P}_n}^h(x_0) = \underset{w^1,\dots,w^n \sim p_{\mathcal{W}}(.)}{Var} \left[\mathfrak{M}_p^h(\tilde{\mathcal{F}}_n, x_0) \right] \tag{6.82}$$

$$= \underset{w^1,\dots,w^n \sim p_{\mathcal{W}}(.)}{\mathbb{E}} \left[\left(\mathfrak{M}_p^h\left(\tilde{\mathcal{F}}_n, x_0\right) - E_{p,\mathcal{P}_n}^h(x_0) \right)^2 \right] . \tag{6.83}$$

Theorem 6.4.19 (Variance of the MFMC estimator)
$\forall x_0 \in \mathcal{X}$,

$$V_{p,\mathcal{P}_n}^h(x_0) \leq \left(\frac{\sigma_{R^h}(x_0)}{\sqrt{p}} + 2C\alpha_{pT}\left(\mathcal{P}_n\right) \right)^2 \tag{6.84}$$

with

$$C = L_\rho \sum_{t=0}^{T-1} \sum_{i=0}^{T-t-1} \left[L_f(1 + L_h) \right]^i . \tag{6.85}$$

Proof - variance

Lemma 6.4.20 (Variance of a sum of random variables)
Let $X_0, \ldots, X_{T-1}$ be T random variables with finite variances $\sigma_0^2, \ldots, \sigma_{T-1}^2$ respectively. Then,

$$Var\left[\sum_{t=0}^{T-1} X_t\right] \leq \left(\sum_{t=0}^{T-1} \sigma_t\right)^2 . \tag{6.86}$$

Proof. The proof is obtained by induction on the number of random variables using the formula

$$Cov(X_i, X_j) \leq \sigma_i \sigma_j \ , \forall i, j \in \{0, \ldots, T-1\} \tag{6.87}$$

which is a straightforward consequence of the Cauchy-Schwarz inequality.

Proof - variance

Definition 6.4.21

Let $x_0 \in \mathcal{X}$. We denote by $\mathfrak{N}_p^h(\tilde{\mathcal{F}}_n, x_0)$ the random variable

$$\mathfrak{N}_p^h\left(\tilde{\mathcal{F}}_n, x_0\right) = \mathfrak{M}_p^h\left(\tilde{\mathcal{F}}_n, x_0\right) - \frac{1}{p}\sum_{i=1}^{p} \mathbb{E}\left[R^h(x_0)|\Omega^{\tau^i}\right]. \tag{6.88}$$

Proof - variance

According to Lemma 6.4.20, we can write

$$V^h_{p,\mathcal{P}_n}(x_0) \leq \left(\sqrt{\underset{w^1,\dots,w^n \sim p_{\mathcal{W}}(.)}{Var} \left[\frac{1}{p} \sum_{i=1}^{p} \mathbb{E}\left[R^h(x_0) | \Omega^{\tau^i} \right] \right]} + \sqrt{\underset{w^1,\dots,w^n \sim p_{\mathcal{W}}(.)}{Var} \left[\mathfrak{N}^h_p\left(\tilde{\mathcal{F}}_n, x_0 \right) \right]} \right)^2 \tag{6.89}$$

Since all the $\left\{ w^{l^i_t} \right\}_{i=1,t=0}^{i=p,t=T-1}$ are i.i.d. according to $p_{\mathcal{W}}(.)$ (cf proof of Theorem 6.4.9), the law of total expectation gives

$$\underset{w^1,\dots,w^n \sim p_{\mathcal{W}}(.)}{Var} \left[\frac{1}{p} \sum_{i=1}^{p} \mathbb{E}\left[R^h(x_0) | \Omega^{\tau^i} \right] \right] = \frac{\sigma^2_{R^h}(x_0)}{p} \ . \tag{6.90}$$

Proof - variance

Now, let us focus on $\underset{w^1,\ldots,w^n \sim p_{\mathcal{W}}(.)}{Var}\left[\mathfrak{N}_p^h(\tilde{\mathcal{F}}_n, x_0)\right]$. By definition, we have

$$\mathfrak{N}_p^h\left(\tilde{\mathcal{F}}_n, x_0\right) = \frac{1}{p}\sum_{i=1}^{p}\left[\sum_{t=0}^{T-1} r^{l_t^i} - \mathbb{E}\left[R^h(x_0)|\Omega^{\tau^i}\right]\right] \quad . \tag{6.91}$$

Then, according to Lemma 6.4.20, we have

$$\begin{aligned}&\underset{w^1,\ldots,w^n \sim p_{\mathcal{W}}(.)}{Var}\left[\mathfrak{N}_p^h\left(\tilde{\mathcal{F}}_n, x_0\right)\right] \\ &\leq \frac{1}{p^2}\left(\sum_{i=1}^{p}\sqrt{\underset{w^1,\ldots,w^n \sim p_{\mathcal{W}}(.)}{Var}\left[\sum_{t=0}^{T-1} r^{l_t^i} - \mathbb{E}\left[R^h(x_0)|\Omega^{\tau^i}\right]\right]}\right)^2\end{aligned} \tag{6.92}$$

Proof - variance

Then, we can write

$$\underset{w^1,\ldots,w^n\sim p_{\mathcal{W}}(.)}{Var}\left[\sum_{t=0}^{T-1} r^{l_t^i} - \mathbb{E}\left[R^h(x_0)|\Omega^{\tau^i}\right]\right]$$

$$\leq \underset{w^1,\ldots,w^n\sim p_{\mathcal{W}}(.)}{\mathbb{E}}\left[\left(\sum_{t=0}^{T-1} r^{l_t^i} - \mathbb{E}\left[R^h(x_0)|\Omega^{\tau^i}\right]\right)^2\right] \tag{6.93}$$

$$\leq \underset{w^1,\ldots,w^n\sim p_{\mathcal{W}}(.)}{\mathbb{E}}\left[\left(a^h\left(\tau^i, x_0\right) - b^h\left(\tau^i, x_0\right)\right)^2\right] = \left(a^h\left(\tau^i, x_0\right) - b^h\left(\tau^i, x_0\right)\right)^2 \tag{6.94}$$

$$\leq 4C^2(\alpha_{pT}\left(\mathcal{P}_n\right))^2\ , \tag{6.95}$$

since $\sum_{t=0}^{T-1} r^{l_t^i}$ and $\mathbb{E}[R^h(x_0)|\Omega^{\tau^i}]$ both belong to the interval $[b^h(\tau^i, x_0), a^h(\tau^i, x_0)]$ whose width is bounded by $2C\alpha_{pT}(\mathcal{P}_n)$ according to Lemma 6.4.17.

Using Equations (6.89), (6.90), (6.92) and (6.95), we have

$$V_{p,\mathcal{P}_n}^h(x_0) \leq \left(\frac{\sigma_{R^h}(x_0)}{\sqrt{p}} + 2C\alpha_{pT}(\mathcal{P}_n)\right)^2 \tag{6.96}$$

which ends the proof. ■

Thx!