Maximizing a Tree Series in the Representation Space

Guillaume Rabusseau, François Denis

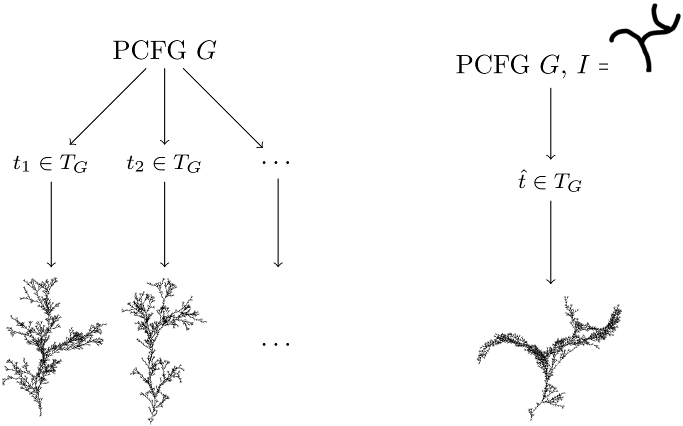
ICGI 2014

September 19, 2014

Overview

Overview

Metropolis Procedural Modeling [Talton et al., 2011]



2 steps:

- Define a *posterior* distribution $p(t|I) \propto \pi(t)L(I|t)$ on T_G
- Find $\hat{t} \in T_G$ maximizing $p(\cdot|I) \Rightarrow$ Metropolis-Hastings

Metropolis-Hastings Algorithm

- $p : \mathcal{X} \to \mathbb{R}_+$ such that $Z = \int_{\mathcal{X}} p(x)\,\mathrm{d}x < \infty$

$\Rightarrow$ $\hat{p} : x \mapsto p(x)\,/Z$ is a probability distribution on $\mathcal{X}$.

Metropolis-Hastings Algorithm

- $p : \mathcal{X} \to \mathbb{R}_+$ such that $Z = \int_{\mathcal{X}} p(x)\,\mathrm{d}x < \infty$

$\Rightarrow$ $\hat{p} : x \mapsto p(x)\,/Z$ is a probability distribution on $\mathcal{X}$.

To sample from $\hat{p}$:

(i) Choose a *jump distribution* $q_x(\cdot)$ (distribution on $\mathcal{X}$ for each $x \in \mathcal{X}$).

(ii) Build a Markov chain in $\mathcal{X}$:

Input : $\mathbf{x_n} \in \mathcal{X}$

Returns : $\mathbf{x_{n+1}} \in \mathcal{X}$

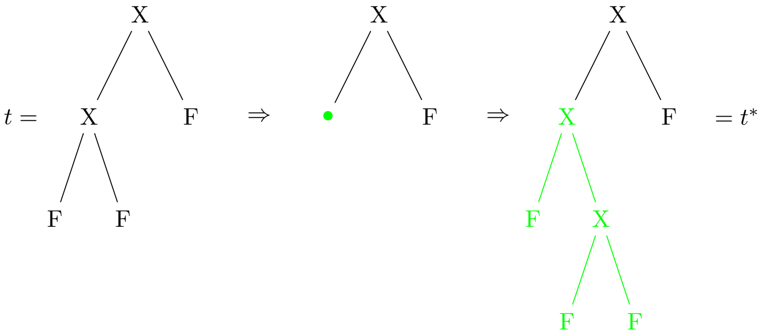
1: Draw a candidate $\mathbf{x}^* \sim q_{\mathbf{x_n}}(\cdot)$

2: Accept $\mathbf{x}^*$ (i.e. $\mathbf{x_{n+1}} \leftarrow \mathbf{x}^*$, otherwise $\mathbf{x_{n+1}} \leftarrow \mathbf{x_n}$) with probability

$$\alpha(\mathbf{x_n}, \mathbf{x}^*) = \min\left\{1, \frac{p(\mathbf{x}^*)q_{\mathbf{x}^*}(\mathbf{x_n})}{p(\mathbf{x_n})q_{\mathbf{x_n}}(\mathbf{x}^*)}\right\}$$

Maximizing $p(\cdot|I)$ with Metropolis-Hastings

- Just need to choose a jump distribution $q_t(\cdot)$:



- And build the Markov chain in T_G :

$$t_0 \xrightarrow{MH} t_1 \xrightarrow{MH} t_2 \xrightarrow{MH} t_3 \xrightarrow{MH} \cdots$$

Overview

Preliminaries: Tree Series

- Set of *tree series* : $\mathbb{R}\langle\langle\mathcal{F}\rangle\rangle = \{\phi : T_{\mathcal{F}} \to \mathbb{R}\}$
- $r \in \mathbb{R}\langle\langle\mathcal{F}\rangle\rangle$ is *rational* $\Leftrightarrow$ r is recognizable by a WTA

Preliminaries: Tree Series

- Set of *tree series* : $\mathbb{R}\langle\langle\mathcal{F}\rangle\rangle = \{\phi : T_{\mathcal{F}} \to \mathbb{R}\}$
- $r \in \mathbb{R}\langle\langle\mathcal{F}\rangle\rangle$ is *rational* $\Leftrightarrow$ r is recognizable by a WTA

Theorem

A tree series $r \in \mathbb{R}\langle\langle\mathcal{F}\rangle\rangle$ is rational if and only if it has a (finite dimensional) linear representation, i.e. there exist (V, μ, λ) s.t. :

- *V is a finite dimensional vector space*
- *$\mu : T_{\mathcal{F}} \to V$ is a linear mapping*
- *$\lambda : V \to \mathbb{R}$ is a linear form s.t. $\forall t \in T_{\mathcal{F}} : r(t) = \lambda(\mu(t))$*

Preliminaries: Tree Series

- Set of *tree series* : $\mathbb{R}\langle\langle\mathcal{F}\rangle\rangle = \{\phi : T_\mathcal{F} \to \mathbb{R}\}$
- $r \in \mathbb{R}\langle\langle\mathcal{F}\rangle\rangle$ is *rational* $\Leftrightarrow$ r is recognizable by a WTA

Theorem

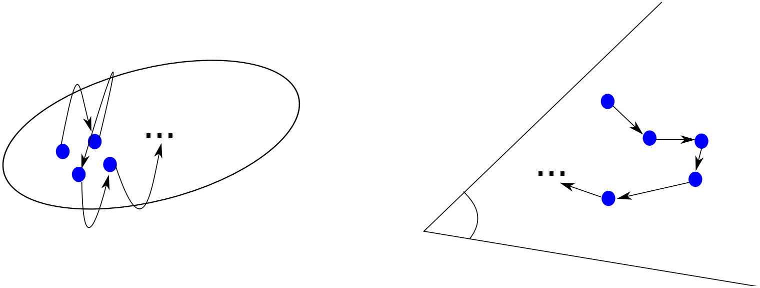
A tree series $r \in \mathbb{R}\langle\langle\mathcal{F}\rangle\rangle$ is rational if and only if it has a (finite dimensional) linear representation, i.e. there exist (V, μ, λ) s.t. :

- *V is a finite dimensional vector space*
- *$\mu : T_\mathcal{F} \to V$ is a linear mapping*
- *$\lambda : V \to \mathbb{R}$ is a linear form s.t. $\forall t \in T_\mathcal{F} : r(t) = \lambda(\mu(t))$*

- (V, μ) is a *linear representation* of $T_\mathcal{F}$.
- V is a *representation space* of $T_\mathcal{F}$.

Problematic

- (Metropolis Procedural Modeling [Talton et al., 2011])
  - Goal: find the tree maximizing $p(\cdot|I) \propto \pi(\cdot)L(I|\cdot)$.
  - Can we use the representation space induced by π ?



- Given a positive tree series $\phi : T_{\mathcal{F}} \to \mathbb{R}$, can we use a representation space of $T_{\mathcal{F}}$ to find

$$\hat{t} = \arg\max_{t \in T_{\mathcal{F}}} \phi(t)$$

Overview

Complexity Study (1)

Definition (Max-RTS)

Input:	A positive rational series $\phi \in \mathbb{R}\langle\langle\mathcal{F}\rangle\rangle$ and $\gamma \in \mathbb{Q}$.
Question:	Is there a tree t such that $\phi(t) \geq \gamma$?

Theorem

- *Max-RTS is undecidable.*
- *With the added constraint that the support of ϕ is finite, Max-RTS is NP-hard.*

Complexity Study (2)

Definition (Ball-RTS)

Input : A linear representation $(\mathbb{R}^n, \mu, \lambda)$ of a positive rational series ϕ, a point $\mathbf{x} \in \mathbb{R}^n$ and $\gamma \in \mathbb{Q}$.

Question : Is there a $t \in supp(\phi)$ s.t. $\|\mu(t) - \mathbf{x}\| \leq \gamma$ (resp. $<, >, \geq$) ?

Theorem

- *Ball-RTS is undecidable.*
- *With the added constraint that the support of ϕ is finite, Ball-RTS is NP-hard.*

Complexity Study (2)

Definition (Ball-RTS)

Input : A linear representation $(\mathbb{R}^n, \mu, \lambda)$ of a positive rational series ϕ, a point $\mathbf{x} \in \mathbb{R}^n$ and $\gamma \in \mathbb{Q}$.

Question : Is there a $t \in supp(\phi)$ s.t. $\|\mu(t) - \mathbf{x}\| \leq \gamma$ (resp. $<$, $>$, $\geq$) ?

Theorem

- *Ball-RTS is undecidable.*
- *With the added constraint that the support of ϕ is finite, Ball-RTS is NP-hard.*

Theorem

Given a linear representation (V, μ) of $T_{\mathcal{F}}$, the injectivity of μ is undecidable.

Maximization with Metropolis-Hastings in $T_{\mathcal{F}}$: MHTF

- Let $\phi : T_{\mathcal{F}} \to \mathbb{R}$ be the positive series to maximize and $\hat{\phi} : t \mapsto \phi(t)/Z$

To sample from $\hat{\phi}(\cdot)$:

(i) Choose a distribution π on $T_{\mathcal{F}}$ and a distribution s_t on $C_{\mathcal{F}}(t)$.

(ii) Build a Markov chain in $T_{\mathcal{F}}$:

Input : $t_n \in T_{\mathcal{F}}$

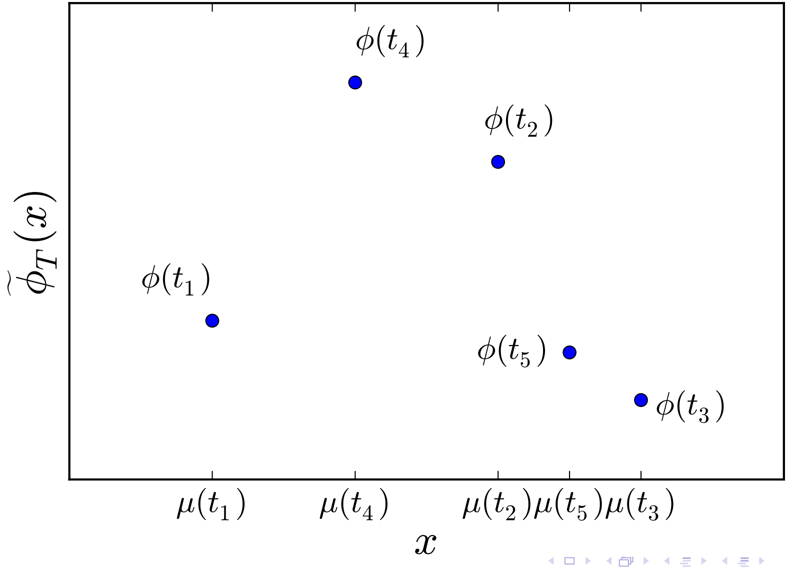
Returns : $t_{n+1} \in T_{\mathcal{F}}$

1: Draw a context $c \in C_{\mathcal{F}}(t_n) \sim s_{t_n}(\cdot)$, and a subtree $\tau \in T_{\mathcal{F}} \sim c^{-1}\pi(\cdot)$

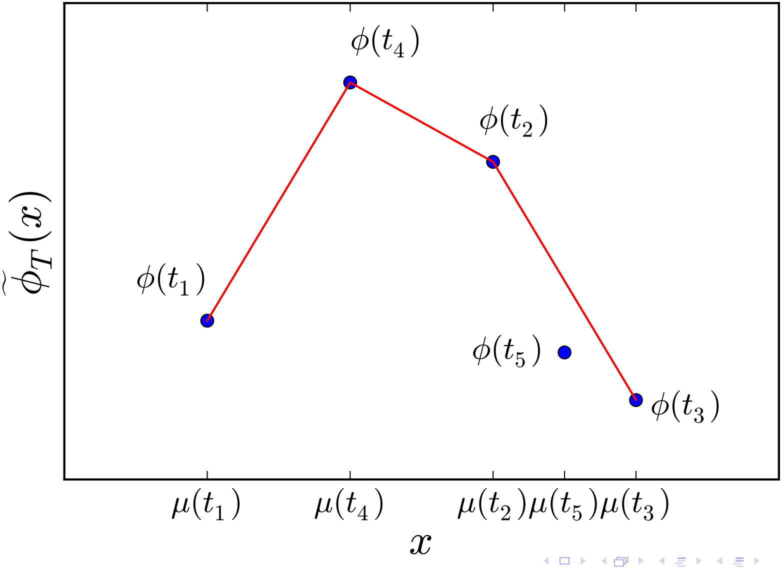
2: $t^* \leftarrow c[\tau]$

3: Accept t^* with probability $\alpha(t_n, t^*) = \min \left\{ 1, \frac{\phi(t^*)s_{t^*}(c)\pi(t_n)}{\phi(t_n)s_{t_n}(c)\pi(t^*)} \right\}$

Extending ϕ to the representation space



Extending ϕ to the representation space



Extending ϕ to the representation space

Let $\phi : T_{\mathcal{F}} \to \mathbb{R}$ be the series to maximize and(V, μ) a linear representation of $T_{\mathcal{F}}$.

Definition

Let $T \subseteq T_{\mathcal{F}}$ be non-empty. Define $\widetilde{\phi}_T$ on $\mathcal{X} = conv(\mu(T))$ by

$$\widetilde{\phi}_T(\mathbf{x}) = \sup_{\substack{n>0,\ \boldsymbol{\alpha}\in[0,1]^n \\ \sum \alpha_i = 1,\ t_1 \cdots t_n \in T}} \left\{ \sum_{i=1}^{n} \alpha_i \phi(t_i) : \mathbf{x} = \sum_{i=1}^{n} \alpha_i \mu(t_i) \right\}$$

for all $\mathbf{x} \in \mathcal{X}$.

Theorem

$\widetilde{\phi}_T$ is a well defined function, continuous on $\mathcal{X}$, and takes its maximum on $\mu(t_{max})$.

Metropolis-Hastings for the distribution $\hat{\phi}_T$ in V

Input : $\mathbf{x_n} \in \mathcal{X}$
Returns : $\mathbf{x_{n+1}} \in \mathcal{X}$

1: Draw a candidate $\mathbf{x}^* \sim q_{\mathbf{x_n}}(\cdot)$
2: Accept $\mathbf{x}^*$ with probability

$$\alpha(\mathbf{x_n}, \mathbf{x}^*) = \min\left\{1, \frac{\hat{\phi}_T(\mathbf{x}^*)q_{\mathbf{x}^*}(\mathbf{x_n})}{\hat{\phi}_T(\mathbf{x_n})q_{\mathbf{x_n}}(\mathbf{x}^*)}\right\} = \min\left\{1, \frac{\widetilde{\phi}_T(\mathbf{x}^*)}{\widetilde{\phi}_T(\mathbf{x_n})}\right\}$$

where $q_{\mathbf{x}}(\cdot)$ is a symmetric and positive jump distribution.

Metropolis-Hastings for the distribution $\hat{\phi}_T$ in V

Input : $\mathbf{x_n} \in \mathcal{X}$

Returns : $\mathbf{x_{n+1}} \in \mathcal{X}$

1: Draw a candidate $\mathbf{x}^* \sim q_{\mathbf{x_n}}(\cdot)$

2: Accept $\mathbf{x}^*$ with probability

$$\alpha(\mathbf{x_n}, \mathbf{x}^*) = \min\left\{1, \frac{\hat{\phi}_T(\mathbf{x}^*) q_{\mathbf{x}^*}(\mathbf{x_n})}{\hat{\phi}_T(\mathbf{x_n}) q_{\mathbf{x_n}}(\mathbf{x}^*)}\right\} = \min\left\{1, \frac{\widetilde{\phi}_T(\mathbf{x}^*)}{\widetilde{\phi}_T(\mathbf{x_n})}\right\}$$

where $q_{\mathbf{x}}(\cdot)$ is a symmetric and positive jump distribution.

Goal : Sample from $\hat{\phi}_{T_\mathcal{F}}$

Idea : Build successive sets of trees $T_1 \subseteq T_2 \subseteq \cdots$ while exploring $\mathcal{X}$.

Adaptive Metropolis-Hastings in V : MHV

Input : $\mathbf{x_n} \in \mathcal{X},\ T_n = \{t_1, \cdots, t_n\} \subseteq T_{\mathcal{F}}$

Returns : $\mathbf{x_{n+1}} \in \mathcal{X},\ T_{n+1} \subseteq T_{\mathcal{F}}$

1: Draw $\mathbf{x}^* \sim q_{\mathbf{x_n}}(\cdot)$

2: Solve $\widetilde{\phi}_{T_n}(\mathbf{x}^*)$: $\mathbf{x}^* = \sum_{i=1}^{n} \alpha_i \mu(t_i)$

3: Draw a tree $t \in T_n$ ($\sim p_\alpha$), a context $c \in C_{\mathcal{F}}(t)$ ($\sim q_t$), and a subtree $\tau \in T_{\mathcal{F}} \sim c^{-1}\pi(\cdot)$

4: $t_{n+1} \leftarrow c[\tau],\ T_{n+1} \leftarrow T_n \cup \{t_{n+1}\}$

5: Accept $\mathbf{x}^*$ with probabilty

$$\alpha(\mathbf{x_n}, \mathbf{x}^*) = \min\left\{1, \frac{\hat{\phi}_{T_n}(\mathbf{x}^*) q_{\mathbf{x}^*}(\mathbf{x_n})}{\hat{\phi}_{T_n}(\mathbf{x_n}) q_{\mathbf{x_n}}(\mathbf{x}^*)}\right\} = \min\left\{1, \frac{\widetilde{\phi}_{T_n}(\mathbf{x}^*)}{\widetilde{\phi}_{T_n}(\mathbf{x_n})}\right\}$$

where $q(\cdot, \cdot)$ is a symmetric and positive jump distribution.

Adaptive Metropolis-Hastings in V : MHV

Input : $\mathbf{x_n} \in \mathcal{X}, \ T_n = \{t_1, \cdots, t_n\} \subseteq T_{\mathcal{F}}$

Returns : $\mathbf{x_{n+1}} \in \mathcal{X}, \ T_{n+1} \subseteq T_{\mathcal{F}}$

1: Draw $\mathbf{x}^* \sim q_{\mathbf{x_n}}(\cdot)$

2: Solve $\widetilde{\phi}_{T_n}(\mathbf{x}^*) : \mathbf{x}^* = \sum_{i=1}^{n} \alpha_i \mu(t_i)$

3: Draw a tree $t \in T_n$ $(\sim p_\alpha)$, a context $c \in C_{\mathcal{F}}(t)$ $(\sim q_t)$, and a subtree $\tau \in T_{\mathcal{F}} \sim c^{-1}\pi(\cdot)$

4: $t_{n+1} \leftarrow c[\tau], \ T_{n+1} \leftarrow T_n \cup \{t_{n+1}\}$

5: Accept $\mathbf{x}^*$ with probabilty

$$\alpha(\mathbf{x_n}, \mathbf{x}^*) = \min\left\{1, \frac{\hat{\phi}_{T_n}(\mathbf{x}^*) q_{\mathbf{x}^*}(\mathbf{x_n})}{\hat{\phi}_{T_n}(\mathbf{x_n}) q_{\mathbf{x_n}}(\mathbf{x}^*)}\right\} = \min\left\{1, \frac{\widetilde{\phi}_{T_n}(\mathbf{x}^*)}{\widetilde{\phi}_{T_n}(\mathbf{x_n})}\right\}$$

Theorem

The distribution of the $\mathbf{x_i}$ *generated by this algorithm converges to* $\hat{\phi}_{T_{\mathcal{F}}}$.

Overview

Setting

Let $\hat{t} \in T_{\mathcal{F}}$, $d(\cdot,\cdot)$ a distance between trees $\phi(\cdot) = \exp(-d(\hat{t},\cdot))$.
Goal: Find t maximizing ϕ (i.e. find $\hat{t}$).

Weighted tree automaton $\mathcal{A} = \langle Q, \mathcal{F}, R \rangle$, where $Q = \{q_1, q_2\}$, $\mathcal{F} = \{f(\cdot,\cdot), a\}$, and R is the set of rules:

$$q_1 \xrightarrow{0.9} f(q_1, q_2) \qquad q_2 \xrightarrow{0.4} f(q_2, q_2)$$

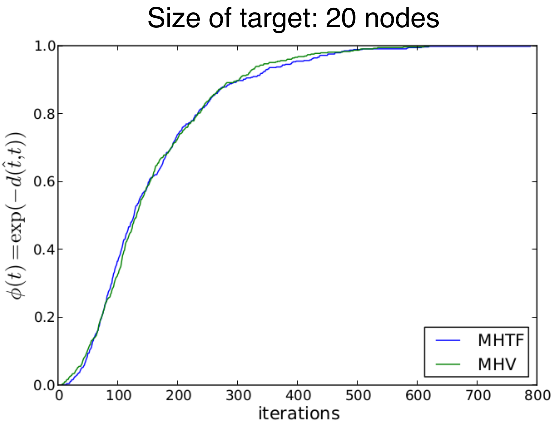
$$q_1 \xrightarrow{0.1} a \qquad q_2 \xrightarrow{0.6} a$$

We run both algorithm

MHTF : MH in $T_{\mathcal{F}}$ using the stochastic series π induced by $\mathcal{A}$ to generate the subtrees.

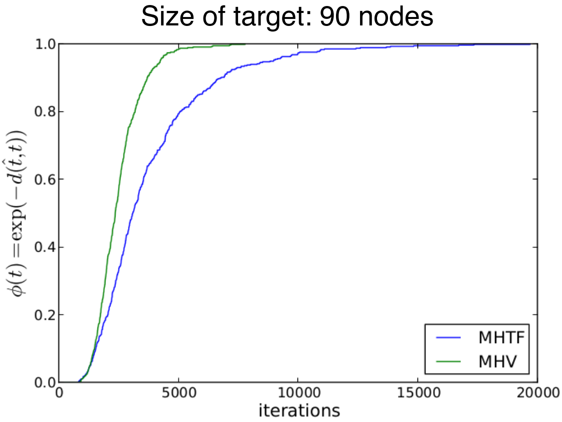
MHV : MH in the representation space induced by $\mathcal{A}$ using π to generate the subtrees.

Results



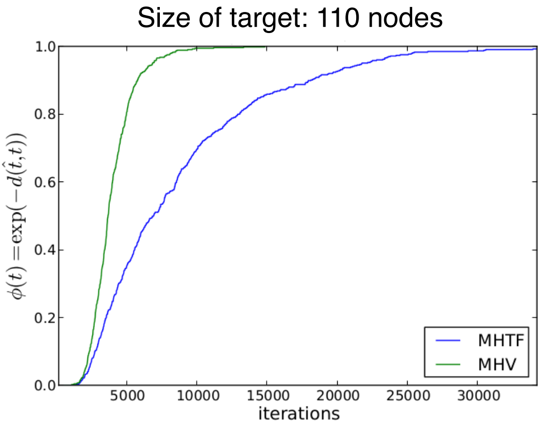
$$\hat{t} = f(f(f(f(f(f(f(f(a,a),a),a),a),f(a,a)),a),a),a)$$

Results



$$\hat{t} = f(f(f(f(f(f(f(f(a, f(f(f(f(f(a, a), a), f(f(f(a, f(a, f(a, f(f(f(f(a, f(a, a)), a), a), a)))), a), a)), a),$$
$$f(a, f(f(f(a, a), f(f(f(a, f(a, a)), a), a)), a)))), a), a), a), f(f(f(a, a), f(f(a, a), f(a,$$
$$f(f(a, a), a)))), a)), a), f(f(f(f(a, a), a), f(f(a, a), a)), a))$$

Results



$$\hat{t} = f(f(f(f(f(f(f(a,a),f(a,f(a,a))),f(f(a,a),f(a,a))),f(f(a,f(f(f(a,a),f(a,f(a,f(a,a)))),a)),f(f(f(f(a,f(f(a,a),f(a,$$
$$f(f(a,f(f(f(a,a),a),f(a,a))),f(f(f(f(a,f(f(f(a,a),a),a)),f(f(a,a),a)),f(a,a)),f(f(a,a),f(a,a))))))),a),f(f(a,a),$$
$$f(f(a,a),f(a,f(a,a))))),a))),f(a,f(a,f(a,a)))),a)$$

Overview

Conclusion and Perspectives

- Contributions
  - Complexity study of problems involving a representation space
  - Algorithm to solve a tree problem in a representation space
  - Convergence guarantees
- Perspectives
  - Quantitative convergence bounds
  - How to choose the representation space?
  - Solve other tree problems in the representation space (e.g. classification)

Berstel, J. and Reutenauer, C. (1982).
Recognizable formal power series on trees.
Theoretical Computer Science, 18(2):115–148.

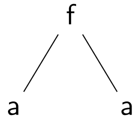
Blondel, V. and Canterini, V. (2003).
Undecidable problems for probabilistic automata of fixed dimension.

Fort, G., Moulines, E., and Priouret, P. (2012).
Convergence of adaptive and interacting Markov chain Monte Carlo algorithms.

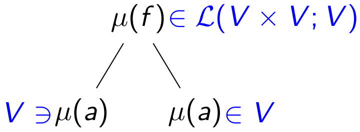
Talton, J. O., Lou, Y., Lesser, S., Duke, J., Mech, R., and Koltun, V. (2011).
Metropolis procedural modeling.

Thank You.

Linear mapping μ



$\Longrightarrow$



$$\mathcal{F} = \{f(\cdot,\cdot), a\}$$

$$\mu(t) = \mu(f)(\mu(a), \mu(a))$$

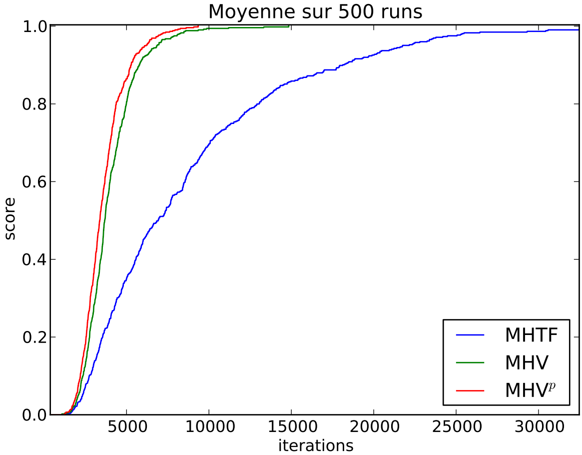
Vitesse d'exécution des algorithmes

- $\hat{t} = f(f(f(f(f(f(f(f(a,a),a),a),a),f(a,a)),a),a),a)$

Algorithme	Nombre moyen d'itérations	Temps moyen d'exécution	Temps moyen par itération
MHTF	159	0.12 s	0.8 ms
MHV	165	0.17 s	1 ms

- $\hat{t} = f(f(f(f(f(f(f(a,f(f(f(f(f(a,a),a),f(f(f(a,f(a,f(a,f(f(f(f(a,f(a,a)),a),a),a)))),a),a)),a),$
$f(a,f(f(f(a,a),f(f(f(a,f(a,a)),a),a)),a)))),a),a),a),f(f(f(a,a),f(f(a,a),f(a,f(f(a,a),a)))),a)),a),$
$f(f(f(f(a,a),a),f(f(a,a),a)),a))$

Algorithme	Nombre moyen d'itérations	Temps moyen d'exécution	Temps moyen par itération
MHTF	3778	7.93 s	2.1 ms
MHV	2595	5.5 s	2.12 ms



$$\hat{t} = f(f(f(f(f(f(a, a), f(a, f(a, a))), f(f(a, a), f(a, a))), f(f(a, f(f(f(a, a), f(a, f(a, f(a, a)))), a)), f(f(f(f(a, f(f(a, a), f(a, f(f(a, f(f(f(a, a), a), f(a, a))), f(f(f(f(a, f(f(f(a, a), a), a)), f(f(a, a), a)), f(a, a)), f(f(a, a), f(a, a))))))), a), f(f(a, a), f(f(a, a), f(a, f(a, a))))), a))), f(a, f(a, f(a, a)))), a)$$