

Transducer Descriptions of DNA Code Properties and Undecidability of Antimorphic Problems

Lila Kari¹ Stavros Konstantinidis² Steffen Kopecki^{1,2}



Descriptive Complexity of Formal Systems 2015
Waterloo, Ontario, Canada — June 25–27

Properties of DNA Libraries

- ▶ DNA alphabet $\Delta = \{A, C, G, T\}$
- ▶ Watson-Crick complement $\delta(A) = T, \delta(T) = A, \delta(C) = G, \delta(G) = C$
- ▶ δ is antimorphic: $\delta(uv) = \delta(v)\delta(u)$
- ▶ DNA library $L \subseteq \Delta^+$ can satisfy all kinds of properties

(a) $\begin{array}{c} \text{GTACCTTCAC} \rightarrow \\ \leftarrow \text{CATGGAAGTG} \end{array}$

(b) $\begin{array}{c} \text{AGTCCAGAAGTTCCTGAATCC} \rightarrow \\ \leftarrow \text{TCTTCAAGGA} \end{array}$

Properties of DNA Libraries

- ▶ DNA alphabet $\Delta = \{A, C, G, T\}$
- ▶ Watson-Crick complement $\delta(A) = T, \delta(T) = A, \delta(C) = G, \delta(G) = C$
- ▶ δ is antimorphic: $\delta(uv) = \delta(v)\delta(u)$
- ▶ DNA library $L \subseteq \Delta^+$ can satisfy all kinds of properties

(a) $\begin{array}{c} \text{GTACCTTCAC} \rightarrow \\ \leftarrow \text{CATGGAAGTG} \end{array}$

(b) $\begin{array}{c} \text{AGTCCAGAAGTTCCTGAATCC} \rightarrow \\ \leftarrow \text{TCTTCAAGGA} \end{array}$

Example properties

- ▶ L is δ -nonoverlapping if $L \cap \delta(L) = \emptyset$.
(a) forbidden, (b) allowed
- ▶ L is δ -compliant if $\forall w \in \delta(L), x, y \in A^*: xwy \in L \implies xy = \varepsilon$.
(a) allowed, (b) forbidden
- ▶ L is strictly δ -compliant if it is δ -nonoverlapping and δ -compliant.
(a) and (b) forbidden

Formal Models

Define a formal model that allows us to (efficiently) solve (some of) the following problems:

Satisfaction given a property P and a (regular) library L :
decide if L satisfies P .

Maximality given a property P and a (regular) library L with $L \in P$:
decide if L is maximal with respect to P .

Construction given a property P and $n \in \mathbb{N}$:
construct a library L such that $|L| \geq n$ and $L \in P$.

Formal Models

Define a formal model that allows us to (efficiently) solve (some of) the following problems:

Satisfaction given a property P and a (regular) library L :
decide if L satisfies P .

Maximality given a property P and a (regular) library L with $L \in P$:
decide if L is maximal with respect to P .

Construction given a property P and $n \in \mathbb{N}$:
construct a library L such that $|L| \geq n$ and $L \in P$.

Two existing models

- ▶ transducer code properties (not suitable for DNA properties)
- ▶ trajectory DNA code properties

Formal Models

Define a formal model that allows us to (efficiently) solve (some of) the following problems:

Satisfaction given a property P and a (regular) library L :
decide if L satisfies P .

Maximality given a property P and a (regular) library L with $L \in P$:
decide if L is maximal with respect to P .

Construction given a property P and $n \in \mathbb{N}$:
construct a library L such that $|L| \geq n$ and $L \in P$.

Two existing models

- ▶ transducer code properties (not suitable for DNA properties)
- ▶ trajectory DNA code properties

Generalization

- ▶ alphabet A with at least two letters
- ▶ (anti-)morphic permutation θ

Classic Transducer Properties

- ▶ A *transducer* $\mathbf{t}: A^* \rightarrow A^*$ is a finite automaton with output.
- ▶ \mathbf{t} is *input-altering* if $w \notin \mathbf{t}(w)$ for all $w \in A^+$.
- ▶ \mathbf{t} is *input-preserving* if $w \in \mathbf{t}(w)$ for all $w \in A^+$.

Definition (Dudzinski, Konstantinidis, 2012)

1.) L has the *input-altering property* described by i.-a. \mathbf{t} if

$$L \cap \mathbf{t}(L) = \emptyset.$$

2.) L has the *input-preserving property* described by i.-p. \mathbf{t} if

$$\forall w \in L: w \notin \mathbf{t}(L \setminus w).$$

Example



- ▶ \mathbf{t}_{pr} is i.-a. and describes the prefix codes.
- ▶ \mathbf{t}_{1s} is i.-p. and describes the 1-substitution error-detection.

θ -Transducer Properties

- ▶ Let $\theta: A^* \rightarrow A^*$ be an (anti-)morphic permutation.
- ▶ \mathbf{t} is *θ -input-altering* if $\theta(w) \notin \mathbf{t}(w)$ for all $w \in A^+$.
- ▶ \mathbf{t} is *θ -input-preserving* if $\theta(w) \in \mathbf{t}(w)$ for all $w \in A^+$.

Definition

1.) L satisfies the *strict θ -transducer property (S-property)* $\mathcal{S}_{\theta, \mathbf{t}}$ if
$$\theta(L) \cap \mathbf{t}(L) = \emptyset.$$

2.) L satisfies the *weak θ -transducer property (W-property)* $\mathcal{W}_{\theta, \mathbf{t}}$ if
$$\forall w \in L: \theta(w) \notin \mathbf{t}(L \setminus w).$$

Remark

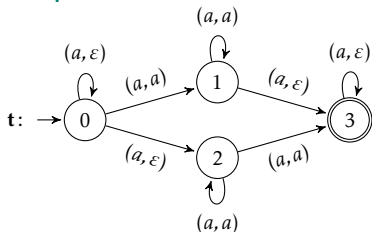
Let $\theta = \text{id}$ be the (morphic) identity.

- 1.) If \mathbf{t} is id-i.-a., then $\mathcal{S}_{\text{id}, \mathbf{t}}$ is the i.-a. transducer property.
- 2.) If \mathbf{t} is id-i.-p., then $\mathcal{W}_{\text{id}, \mathbf{t}}$ is the i.-p. transducer property.

Describing DNA Code Properties

- ▶ L is θ -compliant if $\forall w \in \theta(L), x, y \in A^*: xwy \in L \implies xy = \varepsilon$.

Example

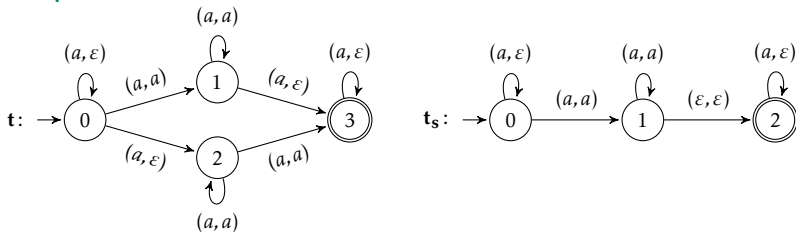


- ▶ \mathbf{t} is θ -i.-a. and $\mathcal{S}_{\theta, \mathbf{t}} = \mathcal{W}_{\theta, \mathbf{t}}$ is the θ -compliant property:
 $\theta(L) \cap \mathbf{t}(L) = \emptyset \iff \forall w \in L: \theta(w) \notin \mathbf{t}(L \setminus w) \iff L$ is θ -compliant

Describing DNA Code Properties

- ▶ L is θ -compliant if $\forall w \in \theta(L), x, y \in A^*: xwy \in L \implies xy = \varepsilon$.
- ▶ L is strictly θ -compliant if $\forall w \in \theta(L), x, y \in A^*: xwy \notin L$.

Example



- ▶ \mathbf{t} is θ -i.-a. and $\mathcal{S}_{\theta, \mathbf{t}} = \mathcal{W}_{\theta, \mathbf{t}}$ is the θ -compliant property:
 $\theta(L) \cap \mathbf{t}(L) = \emptyset \iff \forall w \in L: \theta(w) \notin \mathbf{t}(L \setminus w) \iff L$ is θ -compliant
- ▶ $\mathcal{S}_{\theta, \mathbf{t}_s}$ is the strictly θ -compliant property:
 $\theta(L) \cap \mathbf{t}_s(L) = \emptyset \iff L$ is strictly θ -compliant

First Observations

Proposition

The θ -nonoverlapping property is not describable by any input-preserving transducer.

- ▶ L is θ -nonoverlapping if $L \cap \theta(L) = \emptyset$.

Remark

θ -nonoverlapping is realized by \mathbf{t}_{id} :

$$L \in \mathcal{S}_{\theta, \mathbf{t}_{\text{id}}} \iff \theta(L) \cap \mathbf{t}_{\text{id}}(L) = \emptyset.$$

First Observations

Proposition

The θ -nonoverlapping property is not describable by any input-preserving transducer.

- ▶ L is θ -nonoverlapping if $L \cap \theta(L) = \emptyset$.

Remark

θ -nonoverlapping is realized by \mathbf{t}_{id} :

$$L \in \mathcal{S}_{\theta, \mathbf{t}_{\text{id}}} \iff \theta(L) \cap \mathbf{t}_{\text{id}}(L) = \emptyset.$$

Proposition

The θ -free property is not a θ -transducer property.

- ▶ L is θ -free if and only if $L^2 \cap A^+ \theta(L) A^+ = \emptyset$.

Example

$K = \{\text{ACGT}, \text{CCAC}, \text{GTAA}\}$ is not δ -free as $\text{ACGT} = \delta(\text{ACGT})$ and $\text{CCACGTAA} \in \Delta^+ \text{ACGT} \Delta^+$.

DNA Code Properties via Trajectories

Let $\bar{e}_1, \bar{e}_2 \subseteq \{0, 1\}^*$ be (regular) trajectories.

$$\Phi_{\bar{e}_1, \bar{e}_2}^s(L) = ((L \rightsquigarrow_{\bar{e}_1} A^*) \cap A^+) \sqcup_{\bar{e}_2} A^*$$

\rightsquigarrow scattered deletion

\sqcup shuffle (scattered insertion)

DNA Code Properties via Trajectories

Let $\bar{e}_1, \bar{e}_2 \subseteq \{0, 1\}^*$ be (regular) trajectories.

$$\Phi_{\bar{e}_1, \bar{e}_2}^s(L) = ((L \rightsquigarrow_{\bar{e}_1} A^*) \cap A^+) \sqcup_{\bar{e}_2} A^*$$

\rightsquigarrow scattered deletion

\sqcup shuffle (scattered insertion)

Definition (Domaratzki, 2007)

The *strictly bond-free property described by* (\bar{e}_1, \bar{e}_2) is

$$\mathcal{B}_{\theta}^s(\bar{e}_1, \bar{e}_2) = \{L \subseteq A^* \mid \theta(L) \cap \Phi_{\bar{e}_1, \bar{e}_2}^s(L) = \emptyset\}.$$

N. b.: there is also a *bond-free property*.

Example

$\mathcal{B}_{\theta}^s(1^*0^+1^*, 0^+)$ is the strictly θ -compliant property.

θ -Trajectory vs. θ -Transducer Properties

Theorem

Every regular θ -trajectory property is a θ -transducer \mathcal{S} -property.

Idea: there are transducers that implement all necessary operations

$$\begin{aligned} \mathbf{t}_{\rightsquigarrow_{\bar{e}_1}}(X) &= X \rightsquigarrow_{\bar{e}_1} A^*, & \mathbf{t}_{\sqcup_{\bar{e}_2}}(X) &= X \sqcup_{\bar{e}_2} A^*, & \mathbf{t}_+(X) &= X \cap A^+ \\ \implies \mathbf{t}_{\bar{e}_1, \bar{e}_2}^s &:= \mathbf{t}_{\sqcup_{\bar{e}_2}} \circ \mathbf{t}_+ \circ \mathbf{t}_{\rightsquigarrow_{\bar{e}_1}} &= \Phi_{\bar{e}_1, \bar{e}_2}^s \\ \mathcal{S}_{\theta, \mathbf{t}_{\bar{e}_1, \bar{e}_2}^s} &= \mathcal{B}_{\theta}^s(\bar{e}_1, \bar{e}_2) &= \{L \subseteq A^* \mid \theta(L) \cap \Phi_{\bar{e}_1, \bar{e}_2}^s(L) = \emptyset\} \end{aligned}$$

θ -Trajectory vs. θ -Transducer Properties

Theorem

Every regular θ -trajectory property is a θ -transducer \mathcal{S} -property.

Idea: there are transducers that implement all necessary operations

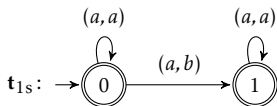
$$\begin{aligned} \mathbf{t}_{\rightsquigarrow_{\bar{e}_1}}(X) &= X \rightsquigarrow_{\bar{e}_1} A^*, & \mathbf{t}_{\sqcup_{\bar{e}_2}}(X) &= X \sqcup_{\bar{e}_2} A^*, & \mathbf{t}_+(X) &= X \cap A^+ \\ \implies \mathbf{t}_{\bar{e}_1, \bar{e}_2}^s &:= \mathbf{t}_{\sqcup_{\bar{e}_2}} \circ \mathbf{t}_+ \circ \mathbf{t}_{\rightsquigarrow_{\bar{e}_1}} &= \Phi_{\bar{e}_1, \bar{e}_2}^s \\ \mathcal{S}_{\theta, \mathbf{t}_{\bar{e}_1, \bar{e}_2}^s} &= \mathcal{B}_{\theta}^s(\bar{e}_1, \bar{e}_2) = \{L \subseteq A^* \mid \theta(L) \cap \Phi_{\bar{e}_1, \bar{e}_2}^s(L) = \emptyset\} \end{aligned}$$

Theorem

\mathcal{H} is a θ -transducer property, but not a θ -trajectory one.

- ▶ $\mathcal{H} = \{L \subseteq A^* \mid \forall u, v \in L: \text{Hamming}(u, \theta(v)) \geq 2\}$

Idea: $\mathcal{S}_{\theta, \mathbf{t}_{1s}}$ is the desired property.



A θ -trajectory property would require “del i ”- and “ins i ”-trajectories. But, “del i ” could combine with “ins j ”.

(Un-)decidability of Satisfaction and Maximality

Given: NFA \mathbf{a} , transducer \mathbf{t} , antimorphic permutation θ

Problem	Property $\mathcal{S}_{\theta,\mathbf{t}}$		Property $\mathcal{W}_{\theta,\mathbf{t}}$	
	general	θ -i.-altering	general	θ -i.-preserving
Satisfaction	decidable* in $\mathcal{O}(\mathbf{t} \mathbf{a} ^2)$		decidable [†]	decidable* in $\mathcal{O}(\mathbf{t} ^2 \mathbf{a} ^4)$
Maximality	undecidable [†]	decidable [†] , PSPACE-hard		

* as in Dudzinski, Konstantinidis, 2012

[†] proven here

(Un-)decidability of Satisfaction and Maximality

Given: NFA \mathbf{a} , transducer \mathbf{t} , antimorphic permutation θ

Problem	Property $\mathcal{S}_{\theta,\mathbf{t}}$		Property $\mathcal{W}_{\theta,\mathbf{t}}$	
	general	θ -i.-altering	general	θ -i.-preserving
Satisfaction	decidable* in $\mathcal{O}(\mathbf{t} \mathbf{a} ^2)$		decidable [†]	decidable* in $\mathcal{O}(\mathbf{t} ^2 \mathbf{a} ^4)$
Maximality	undecidable [†]	decidable [†] , PSPACE-hard		

* as in Dudzinski, Konstantinidis, 2012

† proven here

Remark

Transforming a trajectory property into an \mathcal{S} -properties and deciding satisfaction is asymptotically as fast as deciding the trajectory property directly $\mathcal{O}(|\bar{e}_1||\bar{e}_2||\mathbf{a}|^2) = \mathcal{O}(|\mathbf{t}_{\bar{e}_1, \bar{e}_2}^{\mathbf{s}}||\mathbf{a}|^2)$.

Deciding if L satisfies $\mathcal{W}_{\theta,t}$

Theorem

It is decidable if a regular L satisfies $\mathcal{W}_{\theta,t}$.

Deciding if L satisfies $\mathcal{W}_{\theta,t}$

Theorem

It is decidable if a regular L satisfies $\mathcal{W}_{\theta,t}$.

Lemma

L satisfies $\mathcal{W}_{\theta,t}$ iff $(x, y) \in \mathbf{s} \implies \theta(x) = y$.

$$\blacktriangleright \mathbf{s} = \mathbf{t} \cap (L \times \theta(L)) \subseteq A^* \times A^*$$

Deciding if L satisfies $\mathcal{W}_{\theta,t}$

Theorem

It is decidable if a regular L satisfies $\mathcal{W}_{\theta,t}$.

Lemma

L satisfies $\mathcal{W}_{\theta,t}$ iff $(x, y) \in \mathbf{s} \implies \theta(x) = y$.

$$\blacktriangleright \mathbf{s} = \mathbf{t} \cap (L \times \theta(L)) \subseteq A^* \times A^*$$

Lemma

L satisfies $\mathcal{W}_{\theta,t}$ iff

$$\mathbf{s} \subseteq \bigcup_{\substack{x_1, x_2, x_3 \in A^* \\ |x_1 x_2 x_3| \leq |\mathbf{s}|}} \bigcup_{i \in \mathbb{N}} (x_1 (x_2)^i x_3, \theta(x_1 (x_2)^i x_3)).$$

Deciding if L satisfies $\mathcal{W}_{\theta,t}$

Theorem

It is decidable if a regular L satisfies $\mathcal{W}_{\theta,t}$.

Lemma

L satisfies $\mathcal{W}_{\theta,t}$ iff $(x, y) \in \mathbf{s} \implies \theta(x) = y$.

$$\blacktriangleright \mathbf{s} = \mathbf{t} \cap (L \times \theta(L)) \subseteq A^* \times A^*$$

Lemma

L satisfies $\mathcal{W}_{\theta,t}$ iff

$$\mathbf{s} \subseteq \bigcup_{\substack{x_1, x_2, x_3 \in A^* \\ |x_1 x_2 x_3| \leq |\mathbf{s}|}} \bigcup_{i \in \mathbb{N}} (x_1 (x_2)^i x_3, \theta(x_1 (x_2)^i x_3)).$$

This can be tested by verifying that

- 1.) $\mathbf{s} \subseteq \bigcup_{\substack{x_1, x_2, x_3 \in A^* \\ |x_1 x_2 x_3| \leq |\mathbf{s}|}} (x_1 x_2^* x_3) \times \theta(x_1 x_2^* x_3)$; and
- 2.) $|x| = |y|$ for all $(x, y) \in \mathbf{s}$.

Deciding Maximality

Let $L \in \mathcal{W}_{\theta, \mathbf{t}}$, or $L \in \mathcal{S}_{\theta, \mathbf{t}}$ and \mathbf{t} is θ -input-altering.

Theorem

L is maximal with respect to $\mathcal{W}_{\theta, \mathbf{t}}$ (resp., $\mathcal{S}_{\theta, \mathbf{t}}$) iff

$$L \cup \theta^{-1}(\mathbf{t}(L)) \cup \mathbf{t}^{-1}(\theta(L)) = A^*.$$

Deciding Maximality

Let $L \in \mathcal{W}_{\theta, \mathbf{t}}$, or $L \in \mathcal{S}_{\theta, \mathbf{t}}$ and \mathbf{t} is θ -input-altering.

Theorem

L is maximal with respect to $\mathcal{W}_{\theta, \mathbf{t}}$ (resp., $\mathcal{S}_{\theta, \mathbf{t}}$) iff

$$L \cup \theta^{-1}(\mathbf{t}(L)) \cup \mathbf{t}^{-1}(\theta(L)) = A^*.$$

Corollary

It is PSPACE-hard to decide if L is maximal w. r. t. $\mathcal{W}_{\theta, \mathbf{t}}$ (resp., $\mathcal{S}_{\theta, \mathbf{t}}$).

... because deciding universality of an NFA is PSPACE-hard.

Deciding Maximality

Let $L \in \mathcal{W}_{\theta, \mathbf{t}}$, or $L \in \mathcal{S}_{\theta, \mathbf{t}}$ and \mathbf{t} is θ -input-altering.

Theorem

L is maximal with respect to $\mathcal{W}_{\theta, \mathbf{t}}$ (resp., $\mathcal{S}_{\theta, \mathbf{t}}$) iff

$$L \cup \theta^{-1}(\mathbf{t}(L)) \cup \mathbf{t}^{-1}(\theta(L)) = A^*.$$

Corollary

It is PSPACE-hard to decide if L is maximal w. r. t. $\mathcal{W}_{\theta, \mathbf{t}}$ (resp., $\mathcal{S}_{\theta, \mathbf{t}}$).

... because deciding universality of an NFA is PSPACE-hard.

Note

Let $L \in \mathcal{S}_{\theta, \mathbf{t}}$ and \mathbf{t} is not θ -input-altering. If

$$w \notin L \cup \theta^{-1}(\mathbf{t}(L)) \cup \mathbf{t}^{-1}(\theta(L)),$$

then L may still be maximal provided $\theta(w) \in \mathbf{t}(w)$.

Undecidability of Maximality w. r. t. $\mathcal{S}_{\theta, \mathbf{t}}$

Post correspondence problem (PCP)

Given words $\alpha_1, \alpha_2, \dots, \alpha_\ell \in A^+$ and $\beta_1, \beta_2, \dots, \beta_\ell \in A^+$, decide if there is a non-empty sequence i_1, i_2, \dots, i_n such that

$$\alpha_{i_1} \alpha_{i_2} \cdots \alpha_{i_n} = \beta_{i_1} \beta_{i_2} \cdots \beta_{i_n}.$$

Theorem

It is undecidable if a given transducer is θ -input-preserving.

Idea: define a transducer \mathbf{t} such that

$$\theta(w) \notin \mathbf{t}(w) \iff w = \text{bin}(u \cdot i_n \cdots i_1) \text{ with } u = \alpha_{i_1} \cdots \alpha_{i_n} = \beta_{i_1} \cdots \beta_{i_n}.$$

Undecidability of Maximality w. r. t. $\mathcal{S}_{\theta, \mathbf{t}}$

Post correspondence problem (PCP)

Given words $\alpha_1, \alpha_2, \dots, \alpha_\ell \in A^+$ and $\beta_1, \beta_2, \dots, \beta_\ell \in A^+$, decide if there is a non-empty sequence i_1, i_2, \dots, i_n such that

$$\alpha_{i_1} \alpha_{i_2} \cdots \alpha_{i_n} = \beta_{i_1} \beta_{i_2} \cdots \beta_{i_n}.$$

Theorem

It is undecidable if a given transducer is θ -input-preserving.

Idea: define a transducer \mathbf{t} such that

$$\theta(w) \notin \mathbf{t}(w) \iff w = \text{bin}(u \cdot i_n \cdots i_1) \text{ with } u = \alpha_{i_1} \cdots \alpha_{i_n} = \beta_{i_1} \cdots \beta_{i_n}.$$

Corollary

It is undecidable if \emptyset is maximal with respect to $\mathcal{S}_{\theta, \mathbf{t}}$.

Undecidability of the θ -PCP

θ -Post correspondence problem (θ -PCP)

Given words $\alpha_1, \alpha_2, \dots, \alpha_\ell \in A^+$ and $\beta_1, \beta_2, \dots, \beta_\ell \in A^+$, decide if there is a non-empty sequence i_1, i_2, \dots, i_n such that

$$\alpha_{i_1} \alpha_{i_2} \cdots \alpha_{i_n} = \theta(\beta_{i_1} \beta_{i_2} \cdots \beta_{i_n}).$$

Theorem

The θ -PCP is undecidable.

Undecidability of the θ -PCP

θ -Post correspondence problem (θ -PCP)

Given words $\alpha_1, \alpha_2, \dots, \alpha_\ell \in A^+$ and $\beta_1, \beta_2, \dots, \beta_\ell \in A^+$, decide if there is a non-empty sequence i_1, i_2, \dots, i_n such that

$$\alpha_{i_1} \alpha_{i_2} \cdots \alpha_{i_n} = \theta(\beta_{i_1} \beta_{i_2} \cdots \beta_{i_n}).$$

Theorem

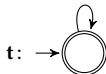
The θ -PCP is undecidable.

Corollary

It is undecidable if a given transducer is θ -input-altering.

Idea:

$$\forall i: (\alpha_i, \theta^2(\beta_i))$$



$$w \in \theta^{-1}(\mathbf{t}(w)) \iff \alpha_{i_1} \cdots \alpha_{i_n} = w = \theta^{-1}(\theta^2(\beta_{i_1} \cdots \beta_{i_n})) = \theta(\beta_{i_1} \cdots \beta_{i_n})$$

Questions?