Edit distance of finite state transducers

C. Aiswarya

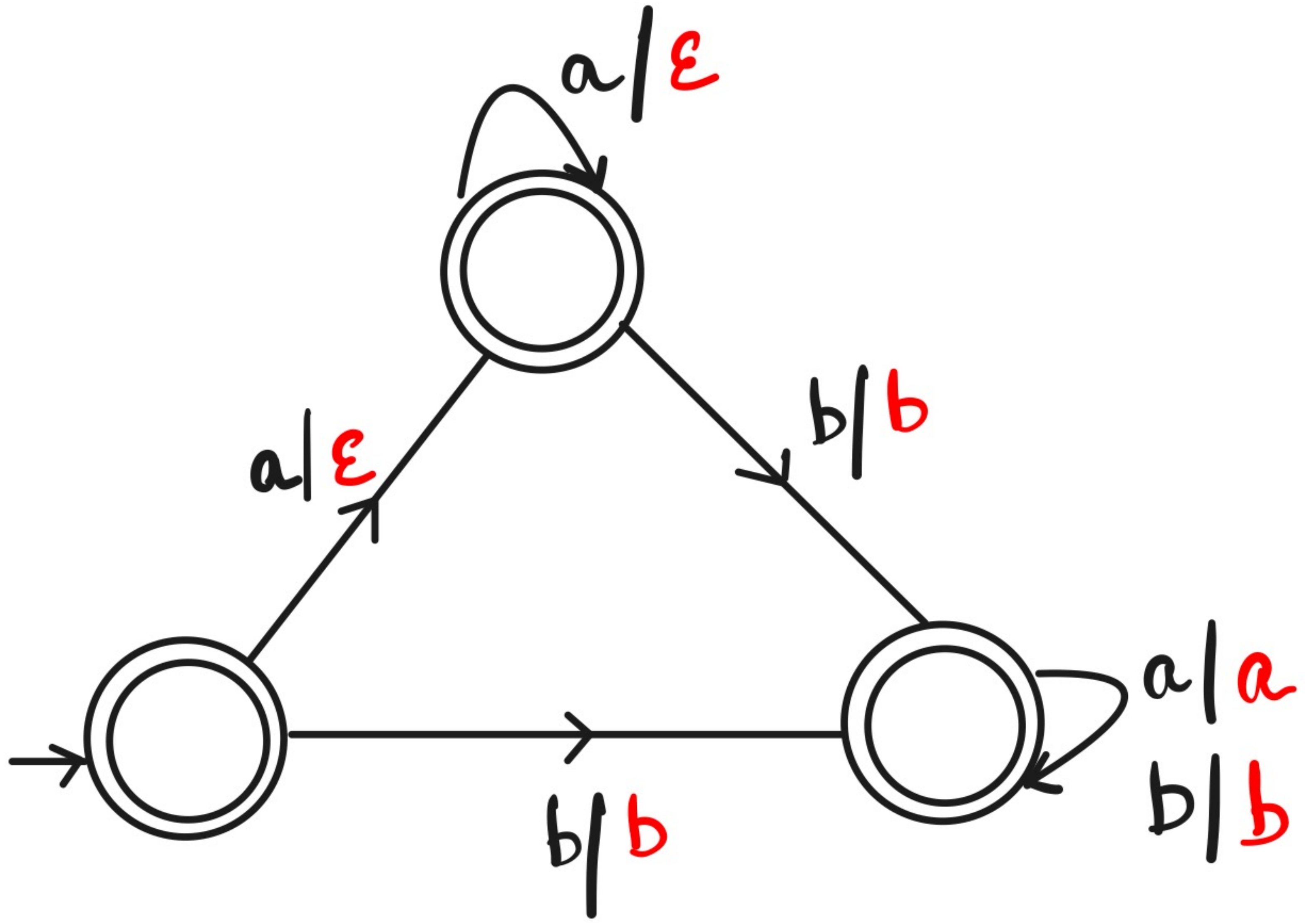
Chennai Mathematical Institute

Joint work with **Amaldev Manuel** and **Saina Sunny** (IIT Goa)

CAALM 03 June 2025

Transducers

- Transducers are finite state automata that produce output words



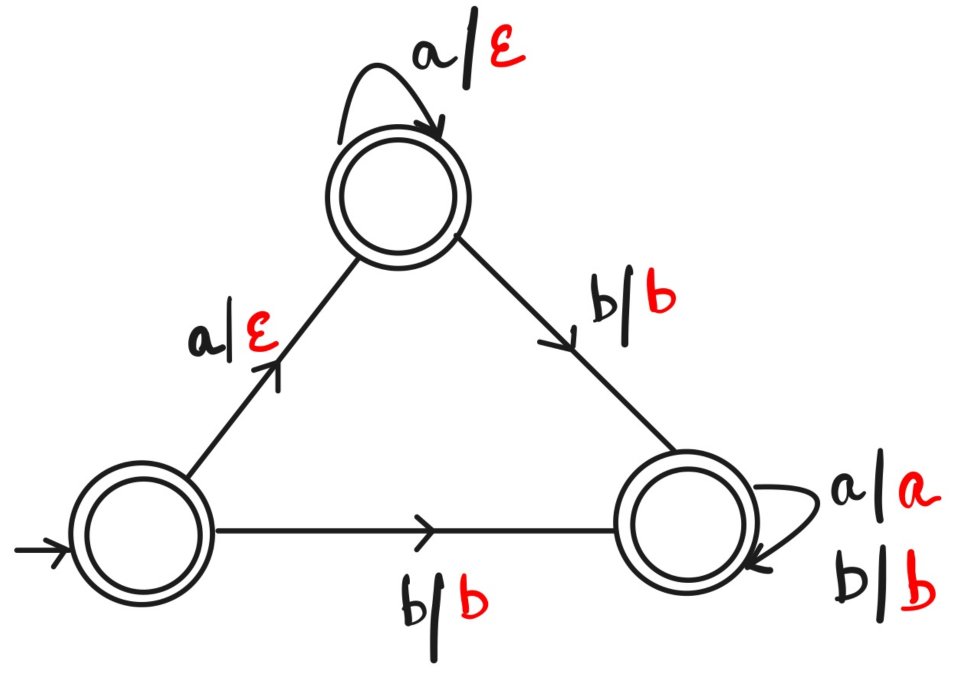
Delete all *a*'s before first *b*

$aaba \to ba$

$bab \to bab$

Transducers

- Transducers are finite state automata that produce output words



Delete all a's before first b

DFA — Sequential Transducers — sequential function

Unambiguous NFA — Unambiguous Transducers — rational function

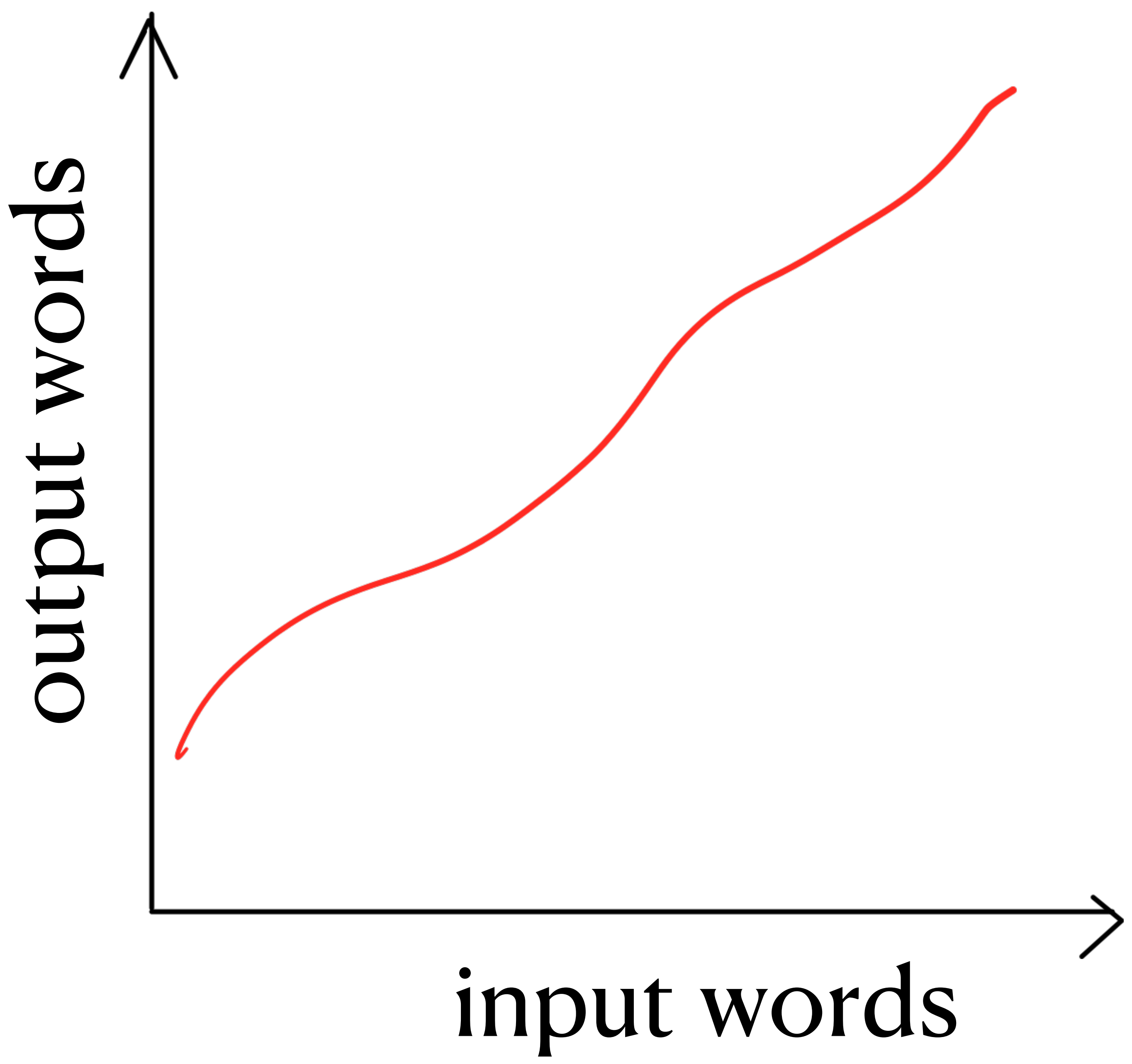
NFA — Rational Transducers — rational relation

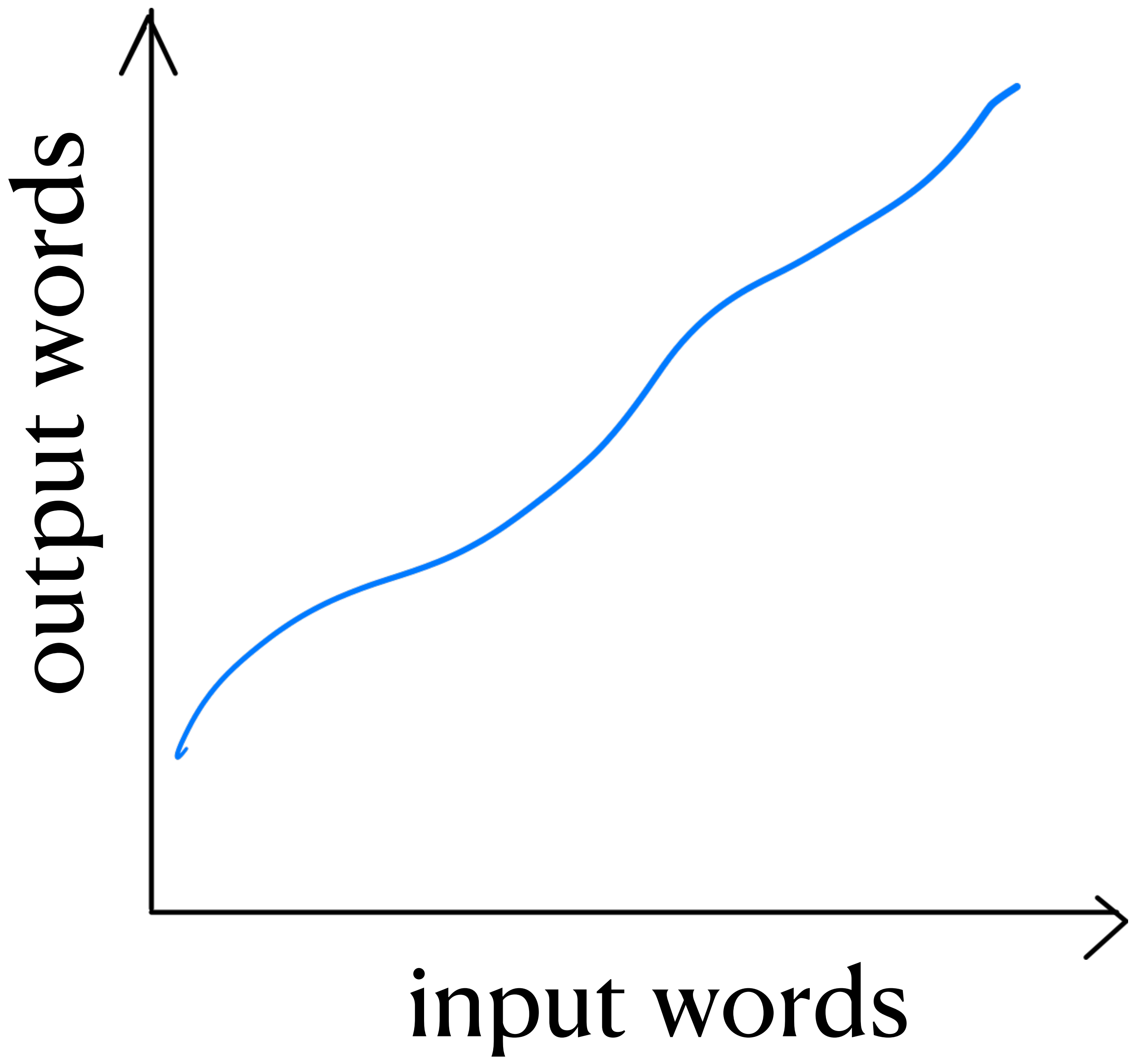
How do we compare transducers?

- Checking equivalence of two transducers
  - decidable for rational functions [Gurari-Ibarra’ 1983],
  - decidable for regular functions [Gurari’1982,Culik-Karhumaki’1987]
  - open for polyregular functions [Bojanczyk’2018]
  - undecidable for rational relations [Fischer-Rosenberg’1968, Griffiths’1968]
- Can we say something meaningful about non-equivalent transducers?

How do we compare transducers?

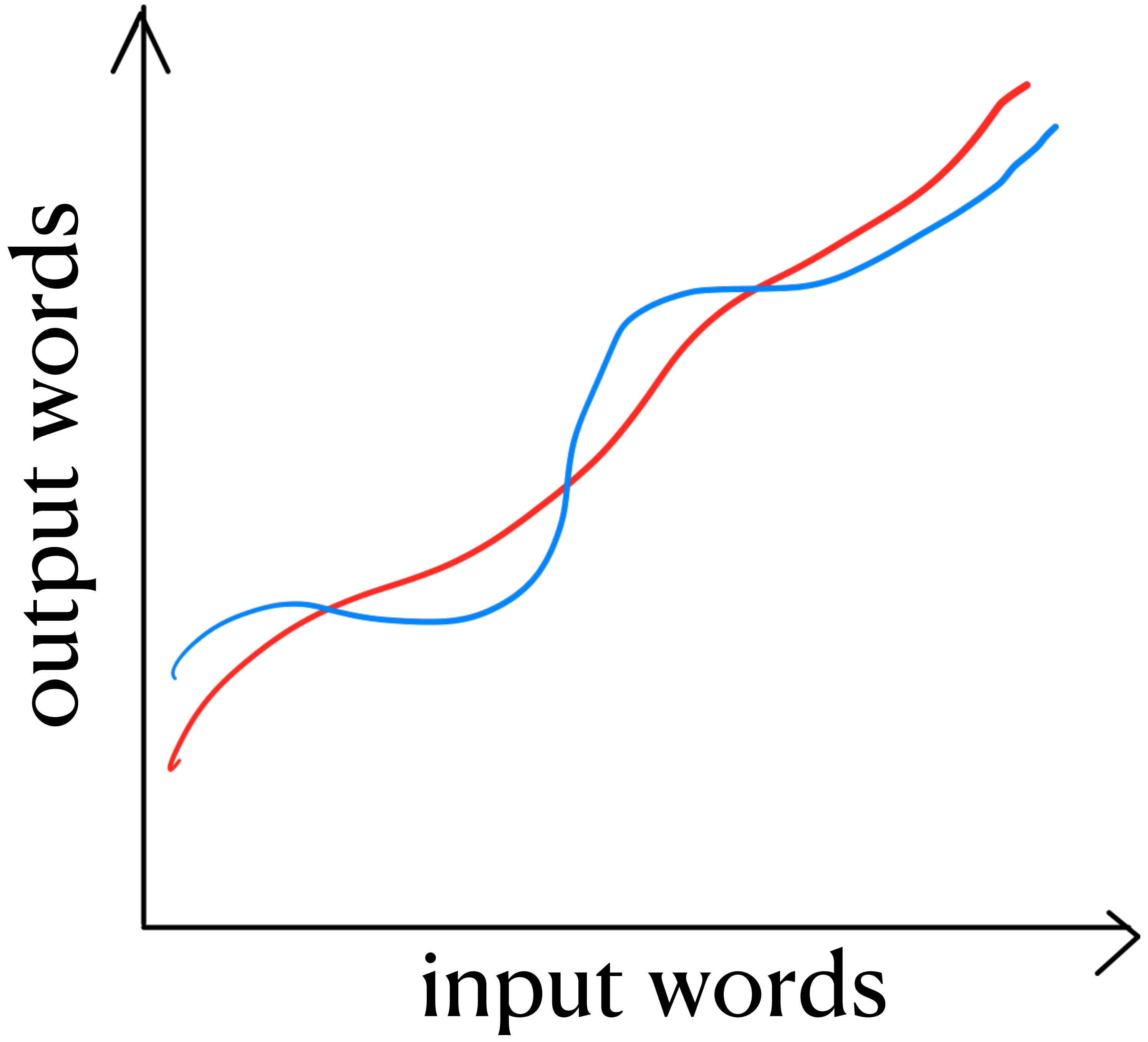
- Functional equivalence (on any input, the respective outputs are “exactly” the same)

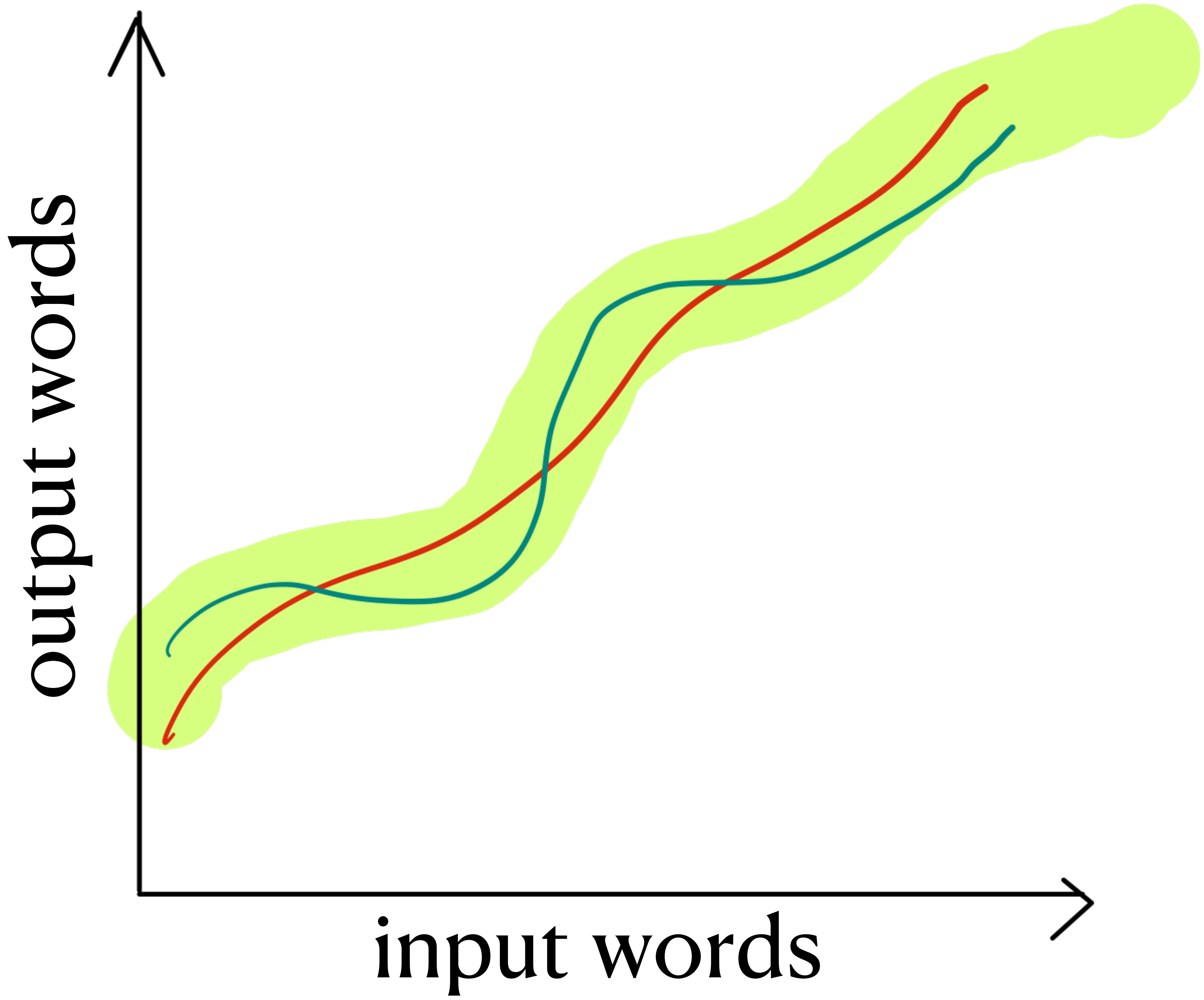




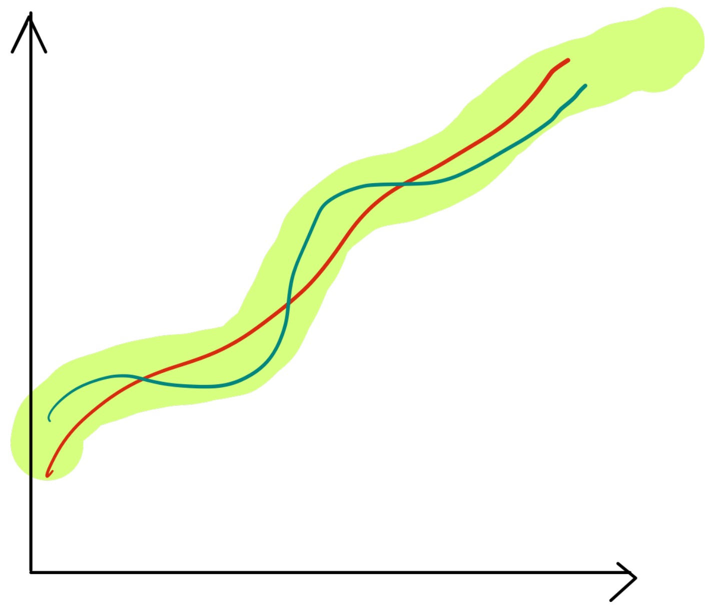
How do we compare transducers?

- Relax it : on any input, the respective outputs are close enough





Metric on transducers



- Let d be a metric on words. Lift it to word-to-word functions (transductions).

$$d(T_1, T_2) = \begin{cases} sup\ \{d(T_1(w), T_2(w)) \mid w \in dom(T_1)\} & \text{if } dom(T_1) = dom(T_2) \\ \infty & otherwise \end{cases}$$

- T_1 and T_2 are close if $d(T_1, T_2)$ is finite.

Edit Distances

- Given a set of edit operations,
    - Ex: insert a letter, delete a letter, or substitute a letter with another

- Edit distances between two words is the minimum number of edits required to convert one to another.

ababa

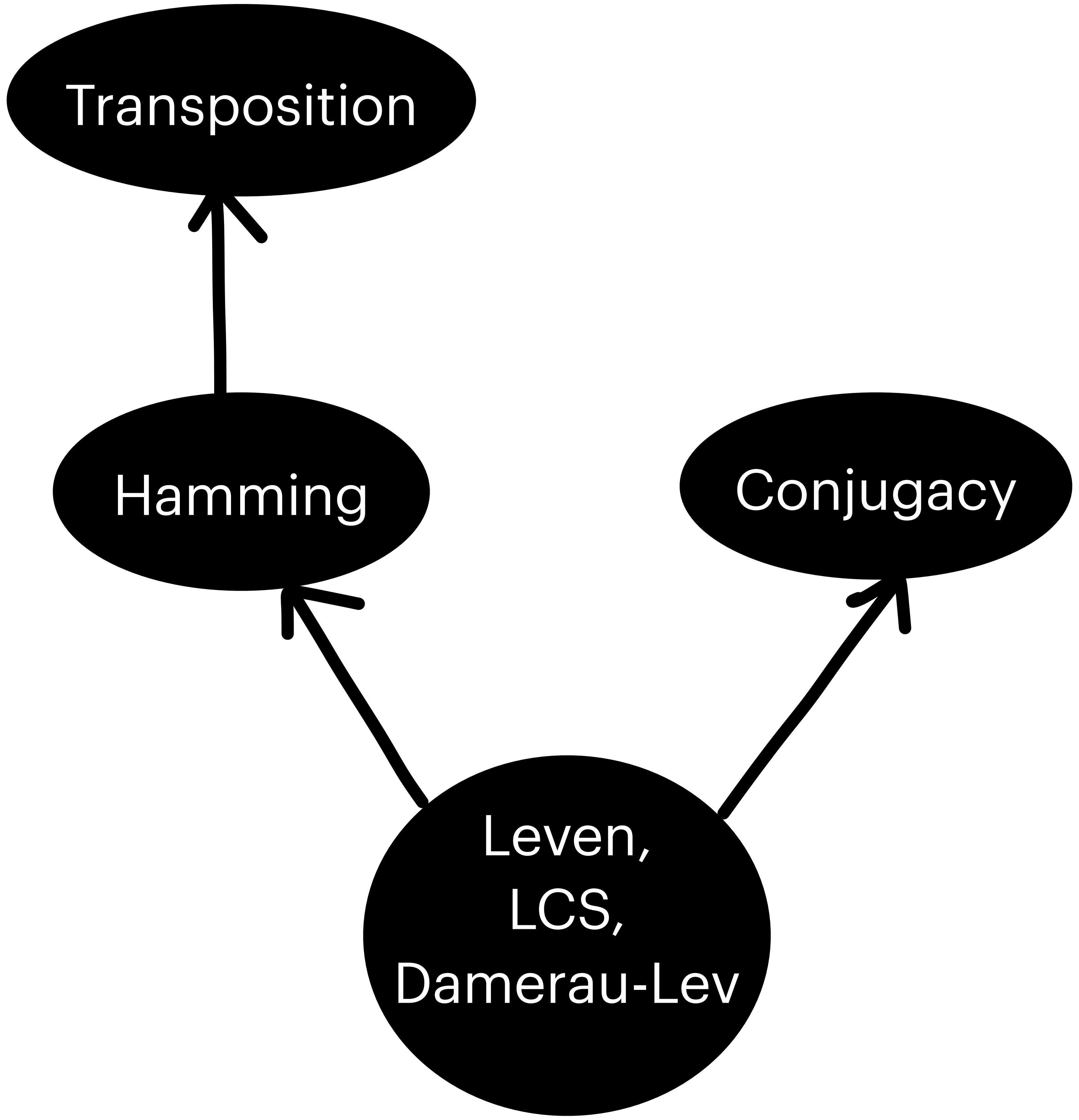
babab

Common Edit distances

Edit Distances	Edit operations
Hamming distance	letter-to-letter substitution
Transposition distance	swapping adjacent letters
Conjugacy distance	left and right cyclic shifts
Levenshtein edit distance	insertion, deletion, substitution
Longest common subsequence	insertion and deletion
Damerau-Levenshtein distance	Insertion,deletion,substitution and adjacent transposition

Edit distances - preorder relation

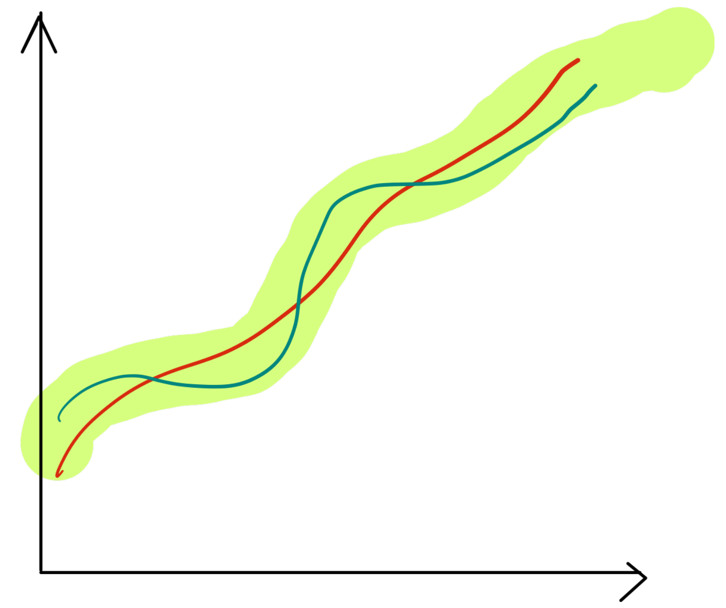
Edit Distances	Edit operations
Hamming distance	letter-to-letter substitution
Transposition distance	swapping adjacent letters
Conjugacy distance	left and right cyclic shifts
Levenshtein edit distance	insertion, deletion, substitution
Longest common subsequence	insertion and deletion
Damerau-Levenshtein distance	Insertion,deletion,substitution and adjacent transposition



Common Edit distances

Edit Distances	Edit operations
Hamming distance	letter-to-letter substitution
Transposition distance	swapping adjacent letters
Conjugacy distance	left and right cyclic shifts
Levenshtein edit distance	insertion, deletion, substitution
Longest common subsequence	insertion and deletion
Damerau-Levenshtein distance	Insertion,deletion,substitution and adjacent transposition
Discrete	$\varnothing$

Metric on transducers



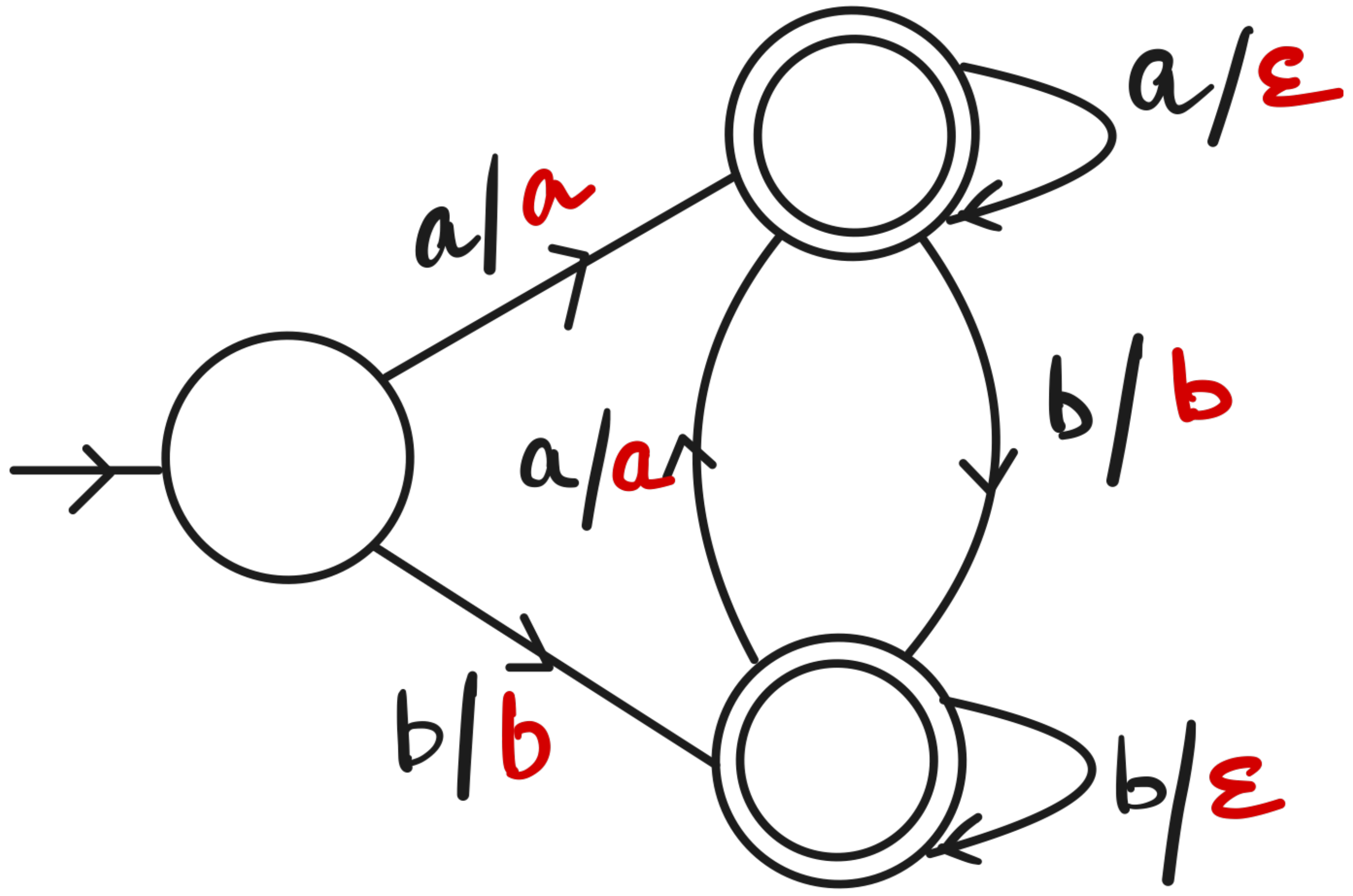
- Let d be a metric on words. Lift it to word-to-word functions (transductions).

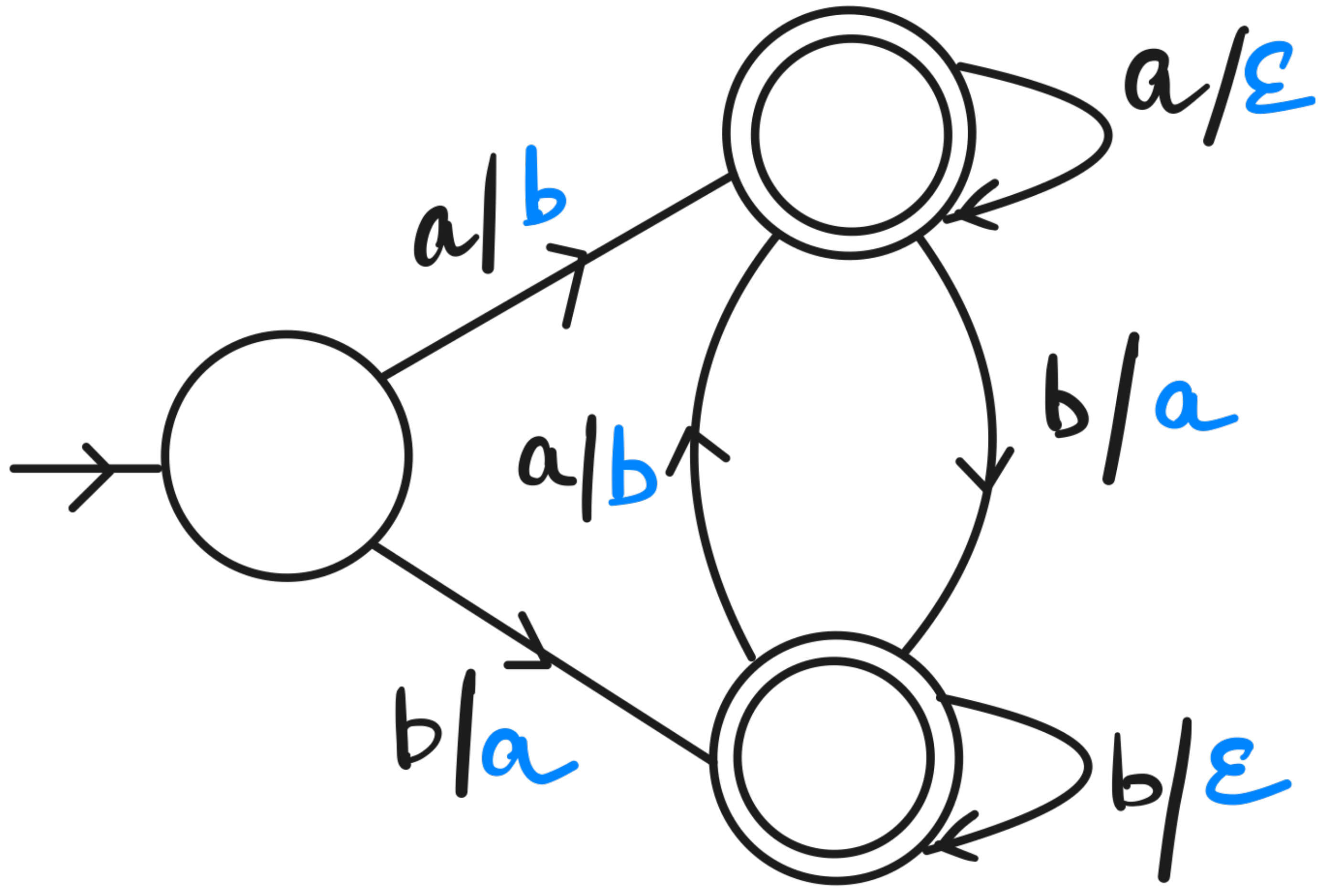
$$d(T_1, T_2) = \begin{cases} sup\ \{d(T_1(w), T_2(w)) \mid w \in dom(T_1)\} & \text{if } dom(T_1) = dom(T_2) \\ \infty & otherwise \end{cases}$$

- T_1 and T_2 are close if $d(T_1, T_2)$ is finite.

Metric on transducers

Example





- For each block of a, output a
- For each block of b, output b

- For each block of a, output b
- For each block of b, output a

$$aaabbabbba \to (ababa, babab)$$

- $d_{lev}(T_1, T_2) = 2$
- $d(T_1, T_2) = \infty$ if only
  - substitutions
  - cyclic shifts
  - adjacent swapping

Metric on transducers

Questions

$$d(T_1, T_2) = \begin{cases} sup\ \{d(T_1(w), T_2(w)) \mid w \in dom(T_1)\} & \text{if } dom(T_1) = dom(T_2) \\ \infty & otherwise \end{cases}$$

- Given T_1, T_2 is $d(T_1, T_2)$ computable? (Distance)
- Given T_1, T_2 is $d(T_1, T_2)$ finite? (Closeness)
- Given T_1, T_2 and $k \in \mathbb{N}$, is $d(T_1, T_2)$ at most k? (k-closeness)

Metric on transducers

Results

Problem	Input	Question
Distance Problem	transducers $\mathcal{T}_1, \mathcal{T}_2$	$d(\mathcal{T}_1, \mathcal{T}_2)$?
Closeness Problem	transducers $\mathcal{T}_1, \mathcal{T}_2$	Is $d(\mathcal{T}_1, \mathcal{T}_2) < \infty$?
k-closeness Problem	integer k, transducers $\mathcal{T}_1, \mathcal{T}_2$	Is $d(\mathcal{T}_1, \mathcal{T}_2) \leq k$?

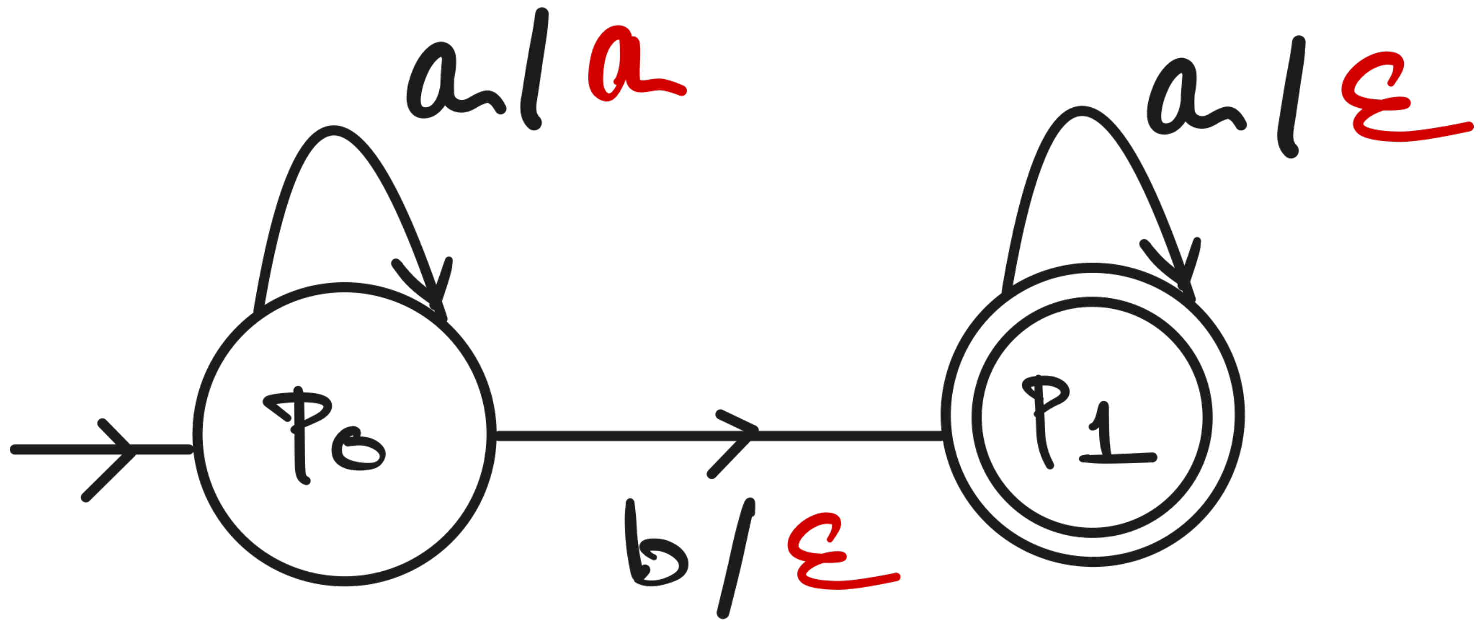
Proposition: Distance is computable iff closeness and k-closeness is decidable for integer-valued metrics

Theorem: Closeness and k-closeness for rational functions are decidable for all metrics $d \in \{d_{lev}, d_{lcs}, d_{damerau}, d_{conj}, d_{ham}, d_{trans}\}$.

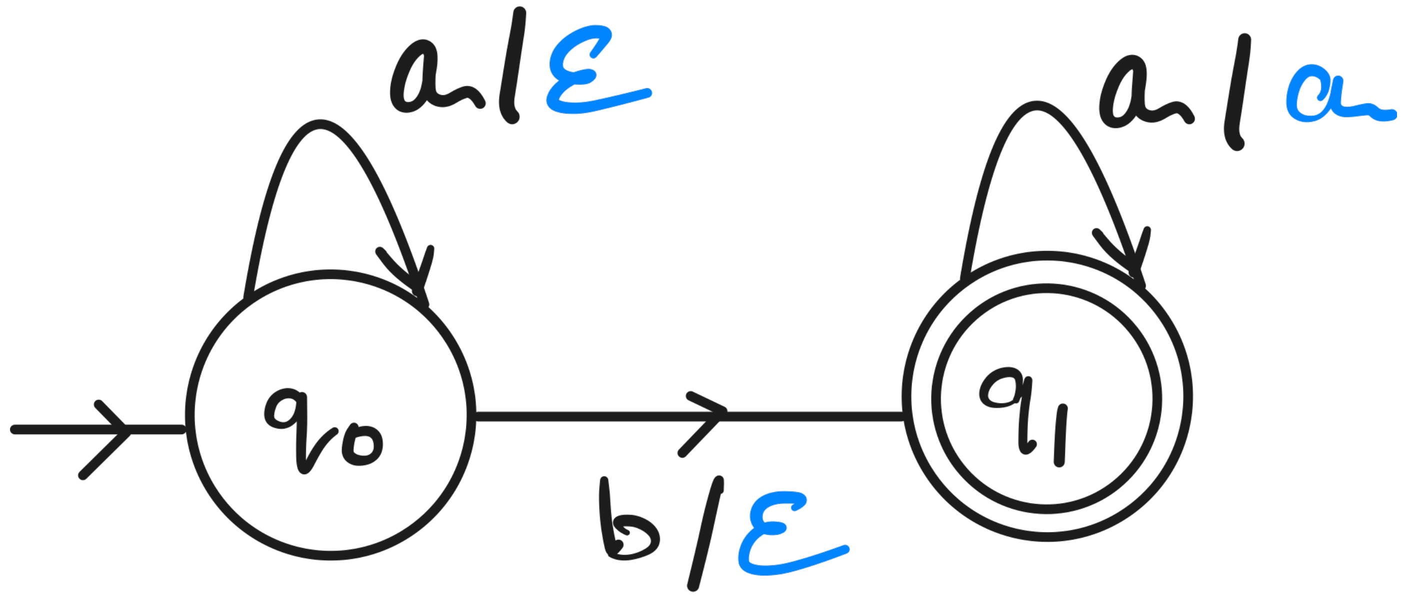
Closeness and k-closeness

- Given transducers T_1, T_2
  - Domain of T_1 and T_2 must be same.
  - Let T be the cartesian product of T_1 and T_2

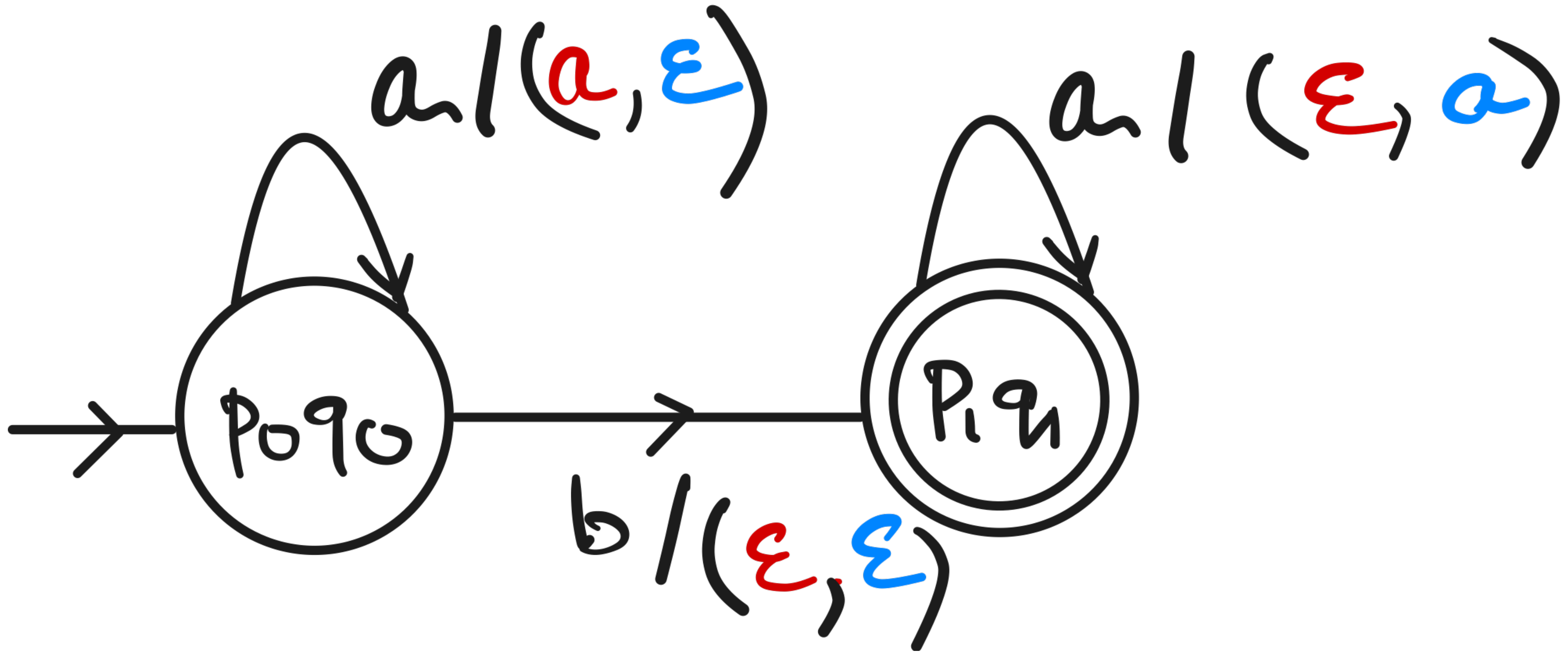
Cartesian product of two transducers



Output a's before b



Output a's after b



$$aaba \rightarrow (a, \epsilon) \cdot (a, \epsilon) \cdot (\epsilon, \epsilon) \cdot (\epsilon, a) = (aa, a)$$

Closeness and k-closeness

- Given transducers T_1, T_2
  - Domain of T_1 and T_2 must be same.
  - Let T be the cartesian product of T_1 and T_2

    — generates set of all pairs of output words of T_1, T_2 on any input
  - Loops of T - must generate output pairs of same length (Close w.r.t. d_{len})

k-closeness
For edit distances

- Given transducers T_1, T_2
  - Domain of T_1 and T_2 must be same.
  - Let T be the cartesian product of T_1 and T_2

    — generates set of all pairs of output words of T_1, T_2 on any input
  - Loops of T - must generate output pairs of same length (Close w.r.t. d_{len})

1. From T, construct an automaton that accepts w if $d(T_1(w), T_2(w)) \leq k$
2. Start with budget k. Non-deterministically do edits, update the budget and residues appropriately. Budget is not allowed to be negative.
3. Check if the language accepted is the domain of T. Yes: k-close; No: not k-close.

Closeness

For edit distances

- Given transducers T_1, T_2
  - Domain of T_1 and T_2 must be same.
  - Let T be the cartesian product of T_1 and T_2

    — generates set of all pairs of output words of T_1, T_2 on any input
  - Loops of T - must generate output pairs of same length (Close w.r.t. d_{len})
  - Loops of T - must generate id pairs? Not necessarily
  - Characterisation based on conjugacy
  - Decidable

Metric on transducers

Results

Problem	Input	Question
Distance Problem	transducers $\mathcal{T}_1, \mathcal{T}_2$	$d(\mathcal{T}_1, \mathcal{T}_2)$?
Closeness Problem	transducers $\mathcal{T}_1, \mathcal{T}_2$	Is $d(\mathcal{T}_1, \mathcal{T}_2) < \infty$?
k-closeness Problem	integer k, transducers $\mathcal{T}_1, \mathcal{T}_2$	Is $d(\mathcal{T}_1, \mathcal{T}_2) \leq k$?

Proposition: Distance is computable iff closeness and k-closeness is decidable for integer-valued metrics

Theorem: Closeness and k-closeness for rational function is decidable for all metrics $d \in \{d_{lev}, d_{lcs}, d_{damerau}, d_{conj}, d_{ham}, d_{trans}\}$.

Related notions and generalisations

Diameter of a Rational Relation

- The diameter of a rational relation R w.r.t. a metric d is the supremum of distance of each pair of words in R

$$dia_d(R) = \sup\{d(u, v) \mid (u, v) \in R\}$$

- Related Work: rational relation with bounded delay [Frougny-Sakarovitch'1991]

Diameter of a Rational Relation

Questions

Problem	Input	Question
Diameter Problem	rational relation R	$dia_d(R)$?
Bounded Diameter Problem	rational relation R	Is $dia_d(R) < \infty$?
k-bounded Diameter Problem	integer k, rational relation R	Is $dia_d(R) \leq k$?

Diameter of a Rational Relation

Results

Problem	Input	Question
Diameter Problem	rational relation R	$dia_d(R)$?
Bounded Diameter Problem	rational relation R	Is $dia_d(R) < \infty$?
k-bounded Diameter Problem	integer k, rational relation R	Is $dia_d(R) \leq k$?

Proposition: Diameter problem of a rational relation is mutually reducible to distance problem of two rational functions

Proposition: Diameter problem of a rational relation is mutually reducible to distance problem of two rational functions

- Distance -> Diameter
    - Given two transducers T_1, T_2, check if their domains are equal
    - $d(T_1, T_2) = dia_d(R)$ where R is the relation generated by cartesian product of T_1 and T_2
- Diameter -> Distance
    - By virtue of [Nivat'1968] theorem

Diameter of a Rational Relation

Results

Problem	Input	Question
Diameter Problem	rational relation R	$dia_d(R)$?
Bounded Diameter Problem	rational relation R	Is $dia_d(R) < \infty$?
k-bounded Diameter Problem	integer k, rational relation R	Is $dia_d(R) \leq k$?

Proposition: Diameter problem of a rational relation is mutually reducible to distance problem of two rational functions

Corollary: All the above problems are decidable for rational relation w.r.t. metrics $d \in \{d_{lev}, d_{lcs}, d_{damerau}, d_{conj}, d_{ham}, d_{trans}\}$

Index of relation in a composition closure

- Index of a rational relation R in the composition closure of S is the smallest integer k such that R is contained in at most k-fold composition of S

$$R \subseteq \bigcup_{0 \leq i \leq k} \underbrace{S \circ S \cdots \circ S}_{i\ times}$$

- Example:

  $\{a, b\}^* \times \{a, b\}^*$

  - S — deletes the first a if exists on any input
  - R_k — deletes first k a's if exist on any input
  - R — delete all a's on any input

  - $Index(R_k, S) = k$
  - $Index(R, S) = \infty$

Index of relation in a composition closure

Questions

Problem	Input	Question
Index Problem	rational relation R, S	Index(R, S)?
Bounded (or Finite) Index Problem	rational relation R, S	Is Index$(R, S) < \infty$?
k-bounded Index Problem	integer k, rational relation R, S	Is Index$(R, S) \leq k$?

Index of relation in a composition closure

Results

Problem	Input	Question
Index Problem	rational relation R, S	$\text{Index}(R, S)$?
Bounded (or Finite) Index Problem	rational relation R, S	Is $\text{Index}(R, S) < \infty$?
k-bounded Index Problem	integer k, rational relation R, S	Is $\text{Index}(R, S) \leq k$?

<u>Lemma</u>: It is undecidable to check if a rational relation has a bounded index in the composition closure of an arbitrary rational relation

Metrizable Relation

Proposition: The index of a rational relation in the composition closure of a d-metrizable relation is computable for $d \in \{d_{len}, d_{lev}, d_{lcs}, d_{dl}, d_h, d_{trans}, d_{conj}\}$

- Graph of a relation S - vertices (words) , edge (between related words in S)
- $d_S(u, v)$ = length of the shortest path between u and v in the graph of S
- S is d- metrizable if d_S is equivalent to metric d up to boundedness.

Index of relation in a composition closure

Results

Problem	Input	Question
Index Problem	rational relation R, S	Index(R, S)?
Bounded (or Finite) Index Problem	rational relation R, S	Is Index$(R, S) < \infty$?
k-bounded Index Problem	integer k, rational relation R, S	Is Index$(R, S) \leq k$?

Lemma: It is undecidable to check if a rational relation has a bounded index in the composition closure of an arbitrary rational relation

Corollary: All the above problems are decidable for rational relation in the composition closure of d- metrizable relation for $d \in \{d_{len}, d_{lev}, d_{lcs}, d_{dl}, d_h, d_{trans}, d_{conj}\}$

Conclusion

- We have defined the following notions
  - Distance between rational functions
  - Diameter of rational relation
  - Index of a rational relation in a composition closure
- All are computable w.r.t. metrics $d \in \{d_{len}, d_{lev}, d_{lcs}, d_{dl}, d_h, d_{trans}, d_{conj}\}$

Thank you


FSTTCS 2025
December 15-20, 2025, Goa India
July 4, 11 : Deadline for submission

References

1. C. Aiswarya, Amaldev Manuel, Saina Sunny. Edit-distance of transducers. ICALP 2024
2. C. Aiswarya, Amaldev Manuel, Saina Sunny. Deciding conjugacy of a rational relation. DLT 2024
3. Christiane Frougny and Jacques Sakarovitch. Rational relations with bounded delay. In STACS 1991,volume 480 of LNCS, pages 50-63, 1991.
4. Christophe Reutenauer and Marcel-Paul Shutzenberger. Minimisation of rational word functions. SIAM Journal on Computing, 20(4):669-685, 1991.
5. Eitan M.Gurari. The equivalence problem for deterministic two-way sequential transducers is decidable. SIAM Journal on Computing, 11(3):448-452, 1982
6. Eitan M. Gurari and Oscar H.Ibarra. A note on finitely-valued and finitely ambiguous transducers. Math.Syst.Theory 16(1):61-66, 1983

References

6. Maurice Nivat. Transduction des langages de Chomsky. PhD Thesis, Annales de l'Institut Fourier, 1968.

7. Mikolaj Bojanczyk. Polyregular functions. CoRR abs/1810.08760 , 2018.

8. Karel Culik, and Juhani Karhumaki. The equivalence problem for single-valued two-way transducers (on NPDTOL Languages) is decidable. SIAM Journal on Computing, 1987.

9. Roger C Lyndon, Marcel-Paul Schützenberger, et al. The equation $a^M = b^N c^P$ in a free group. Michigan Math. J, 9(4):289-298, 1962.

10. Timothy V.Griffiths. The unsolvability of the equivalence problem for lambda-free nondeterministic generalized machines. J. ACM 15(3):409-413,1968

11. Patrick C. Fischer, and Arnold L. Rosenberg. Multitape one-way non writing automata, JCSS 2(1), 88-101,1968