# Sobre a minimização de transdutores sequenciais

**Rodrigo de Souza**[1*]

[1]DEINFO – Universidade Federal Rural de Pernambuco (UFRPE)
Dois Irmãos – 52.171-900 – Recife – PE – Brazil

rodrigo.npmsouza@ufrpe.br

***Abstract.*** *We study the minimization of sequential transducers with outputs in structures which we call monoids with gcd ("greatest common divisor"). We prove the existence of a minimal sequential transducer for a sequential function* $\Sigma^* \to \mathbb{M}$*, where* $\mathbb{M}$ *is a cancellative monoid such that every non-empty subset has a single gcd. This result includes the minimization for outputs in a free monoid or in the additive monoid of the non-negative real numbers, among others. Next, we characterize the sequential functions* $\Sigma^* \to \mathbb{M}$*, where* $\mathbb{M}$ *is a cancellative monoid with gcd, by using the right congruence of a function.*

***Resumo.*** *Estudamos a minimização de transdutores sequenciais utilizando para as saídas uma família de monóides que chamamos de monóides com* $\mathrm{mdc}$*. Provamos a existência de um transdutor minimal para funções sequenciais* $\Sigma^* \to \mathbb{M}$*, onde* $\mathbb{M}$ *é um monóide cancelativo com* $\mathrm{mdc}$ *único. Esse resultado inclui diversos monóides de interesse, como os monóides livres, e o monóide aditivo dos números reais não-negativos. Também apresentamos uma caracterização das funções sequenciais* $\Sigma^* \to \mathbb{M}$*, onde* $\mathbb{M}$ *é um monóide cancelativo com* $\mathrm{mdc}$*, utilizando a congruência à direita de uma função.*

## 1. Introdução

Os *autômatos finitos* são tão essenciais na Ciência da Computação quanto os espaços topológicos na Análise e os grafos na Combinatória, por representarem a noção de *computação* em sua forma mais límpida. Em consequência, a disciplina hoje denominada Teoria dos Autômatos firmou-se como um dos pilares da Ciência da Computação.

Em sua definição mais simples, *transdutores*[1] são uma extensão do modelo de autômato no qual as transições são rotuladas por um par de palavras, uma dita *de entrada* e outra *de saída*. Ao invés de reconhecer linguagens, os transdutores realizam *relações* entre linguagens, chamadas *transduções* em algumas referências (especialmente americanas), ou ainda *relações racionais* (especialmente pelos autores filiados à escola francesa). A definição formal repousa no produto cartesiano de dois monóides livres $\Sigma^*$ e $\Gamma^*$, $\Sigma^* \times \Gamma^*$, um monóide cuja operação é a aplicação das concatenaçoes de $\Sigma^*$ e $\Gamma^*$ nas coordenadas respectivas. De forma mais geral, um monóide qualquer, como os números reais com a operação de soma, pode ser utilizado no lugar do monóide de saída $\Gamma^*$.

[*]Este trabalho é apoiado pelo projeto *Problemas estruturais em modelos formais de Computação*, Edital MCTI/CNPQ/Universal 14/2014 (Processo 459957/2014-7).

[1]Por limitação de espaço, não apresentamos neste resumo definições e notações básicas sobre autômatos e transdutores. Esse material pode ser consultado no livro de Jacques Sakarovitch [Sakarovitch 2009].

Versões restritas de transdutores, as chamadas máquinas de Mealy e Moore, estão entre os primeiros tipos de autômatos da literatura, tendo sido propostos na década de 1960. Desde então, os transdutores se firmaram como uma das pedras angulares da Teoria dos Autômatos, citando aqui o compêndio clássico de Samuel Eilenberg [Eilenberg 1974]. Transdutores também tem muitas aplicações práticas. Por exemplo, algoritmos para determinização, minimização e composição de transdutores são extensivamente utilizados em sistemas de tratamento de linguagem natural [Benesty et al. 2008]. Um aspecto importante dessas ferramentas é sua sustentação em propriedades profundas de transdutores cujo estudo remonta à década de 1970 [Choffrut 1979] mas que foram desde então revisitadas através de novas abordagens [Béal et al. 2003, Choffrut 2003]. Um panorama atualizado da extensão da teoria dos transdutores pode ser consultado em [Sakarovitch 2009].

As relações racionais que são funções parciais – ou *funções racionais* – tiveram uma posição predominante nesse contexto devido a propriedades notáveis dessa família (decidibilidade, não-ambiguidade), e em particular devido ao interesse teórico e prático envolvido no conceito de determinismo (na entrada). Contrariamente ao caso clássico dos autômatos sobre um monóide livre, nem todo transdutor funcional (= realiza uma função) pode ser determinizado; a caracterização das funções racionais que admitem uma leitura determinística, ou *sequencial*, em nossa terminologia, é um resultado profundo, obtido por Choffrut em 1979 [Choffrut 1979]. Aquelas que podem ser realizadas por um *transdutor sequencial*[2] – as *funções sequenciais* – são importantes em aplicações práticas.

O presente resumo aborda o conceito de *minimização de transdutores sequenciais*, que estende os resultados clássicos sobre autômatos determinísticos. O estudo de um transdutor minimal que realiza uma função sequencial aparece implicitamente em trabalhos da década de 1970 de Choffrut [Choffrut 1979], e explicitamente em trabalhos posteriores de Reutenauer [Reutenauer 1990] e Choffrut [Choffrut 2003]. Nesses dois últimos, são consideradas funções com saídas em um monóide livre. Ao mesmo tempo, um conceito de transdutor minimal para funções sequenciais com saídas em outros tipos de monóides, como o monóide dos números reais com a operação de adição, também foi estudado por Mohri e colaboradores, em conexão com aplicações em tratamento de linguagem natural [Mohri 2000]. A abordagem ao assunto não é uniforme na literatura: de um lado, constituiu-se uma noção de congruência sintática para funções, semelhante à congruência de Nerode para linguagens, que permite caracterizar algebricamente o transdutor minimal para uma função *com saídas em um monóide livre*. Como ocorre com os autômatos, essa máquina minimal tem o menor número de estados dentre todas aquelas equivalentes, o que pode ser observado através de *morfismos entre transdutores*. Esse tratamento, utilizado por Eilenberg no estudo da minimização de autômatos, aparece superficialmente no trabalho de Reutenauer, e explicitamente no trabalho de Choffrut de 2003. Nessa abordagem, o primeiro passo é definir o transdutor seqüencial minimal associado a uma função, de maneira puramente algébrica. Em seguida, utilizando morfismos de transdutores sequenciais, demonstra-se a minimalidade desse transdutor. No trabalho de Mohri, a abordagem é algorítmica: o objetivo é descrever a construção de um transdutor minimal, tanto para emissões em um monóide livre, quanto em um monóide numérico.

Nossa intenção é apresentar um formalismo permitindo descrever o conceito de

[2]Preferimos o termo *sequencial* a *subsequencial*, usado originalmente. Uma discussão sobre a terminologia (e algoritmos para decidir a sequencialidade) é feita em [Lombardy and Sakarovitch 2006].

autômato minimal de forma unificada, considerando em uma única linguagem diversos monóides já abordados na literatura. Para tanto, definimos uma família de monóides, que chamamos de *monóides com* $\mathrm{mdc}$, que abstrai as propriedades essenciais do monóide de saída para a definição de um transdutor minimal. Nas próximas seções, esboçamos nossos principais resultados, e fazemos referência à nossa dissertação de mestrado para uma discussão mais extensa sobre o assunto e as demonstrações completas [de Souza 2004].

## 2. Transdutores sequenciais minimais

Seja $\mathrm{M}$ um monóide. Dados elementos $a$ e $b$ em $\mathrm{M}$, escrevemos $a|b$ se existir um elemento $b'$ em $\mathrm{M}$ tal que $b = ab'$. Essa relação de "divisibilidade"é transitiva, reflexiva, mas não necessariamente anti-simétrica (considere o caso de um grupo). Dado um subconjunto não-vazio $S$ de $\mathrm{M}$, dizemos que um elemento $m \in \mathrm{M}$ é um *máximo divisor comum*, ou $\mathrm{mdc}$, de $S$, se $m|a$, para todo $a \in S$, e para cada $n \in \mathrm{M}$, se $n|a$, para todo $a \in S$, então $n|m$. Dizemos que $\mathrm{M}$ é um monóide com $\mathrm{mdc}$ se todo subconjunto não-vazio de $\mathrm{M}$ tiver pelo menos um $\mathrm{mdc}$. Interessa-nos o caso de o $\mathrm{mdc}$ de todo subconjunto não-vazio de $\mathrm{M}$ ser único. Dizemos nesse caso que $\mathrm{M}$ é um *monóide com* $\mathrm{mdc}$ *único*. Exemplos dessa estrutura são: monóides livres (o $\mathrm{mdc}$ é a operação de maior prefixo comum de uma linguagem); monóides livres comutativos; o monóide $(\mathbb{Z}, \cdot)$ dos inteiros positivos, com a operação de produto (com o $\mathrm{mdc}$ usual); o monóide $(\mathbb{R}_+, +)$ dos números reais não-negativos, com a operação de adição (o $\mathrm{mdc}$ é o ínfimo de um conjunto). Intuitivamente, o $\mathrm{mdc}$ abstrai a extração do maior prefixo comum de uma linguagem, operação essencial na definição do transdutor minimal com emissões em um monóide livre.

O escopo de nossa definição de transdutor minimal é o das funções sequenciais $f : \Sigma^* \to \mathrm{M}$, onde $\Sigma$ é um alfabeto, e $\mathrm{M}$ é um monóide cancelativo com $\mathrm{mdc}$ único. A teoria aqui desenvolvida é então uma extensão do trabalho de Choffrut de 2003, no qual o monóide de saída é sempre um monóide livre [Choffrut 2003]. Ademais, Choffrut define o transdutor minimal utilizando a congruência sintática de uma função. Fazemos uma construção direta do transdutor minimal a partir de $f$, seguindo as ideias desenvolvidas nas seções III.3 e XII.5 do livro de Eilenberg [Eilenberg 1974].

Em linhas gerais, a construção do transdutor sequencial minimal para $f$, que denotamos $\mathcal{T}_f$, parte do conjunto $Q^0$ de todas as funções parciais não-vazias $g : \Sigma^* \to \mathrm{M}$ tais que $\mathrm{mdc}\, g = 1$ (onde $1$ é a unidade do monóide; para o monóide livre, é a palavra vazia). Esse conjunto é análogo ao conjunto potência de $\Sigma^*$ na definição do autômato minimal para uma linguagem sobre $\Sigma$ (mas note a restrição imposta sobre o $\mathrm{mdc}$). De forma puramente algébrica, definimos então uma ação $\cdot$ de $\Sigma^*$ sobre $Q^0$, e uma emissão $*$ de $\Sigma^*$ sobre $Q^0$ em $\mathrm{M}$ (que correspondem às regras de transição e emissão do transdutor). As restrições de $\cdot$ e $*$ a uma parte especial de $Q^0$ constituem o transdutor $\mathcal{T}_f$. A propriedade mais relevante dessa construção, que garante a minimalidade de $\mathcal{T}_f$, é:

**Teorema 2.1** *Seja $f : \Sigma^* \to \mathrm{M}$ uma função sequencial, onde $\mathrm{M}$ é um monóide cancelativo com $\mathrm{mdc}$ único. Para todo transdutor sequencial bi-acessível $\mathcal{T}$ que realiza $f$, existe um morfismo próprio $\varphi : \mathcal{T} \to \mathcal{T}_f$.*

Como $\varphi : \mathcal{T} \to \mathcal{T}_f$ é uma função sobrejetora, $n_f \leq n$, onde $n_f$ e $n$ são os números de estados de $\mathcal{T}_f$ e $\mathcal{T}$, respectivamente. Essa propriedade não apenas mostra que $\mathcal{T}_f$ é finito, se $f$ for sequencial, mas também que $\mathcal{T}_f$ tem o menor número de estados dentre todos os transdutores que realizam $f$. Ademais, é fácil ver que, se $\mathcal{T}$ tiver o mesmo número de estados de $\mathcal{T}_f$ e for prefixo, então $\mathcal{T}$ e $\mathcal{T}_f$ são isomorfos.

## 3. Congruência de uma função e redução

Nosso estudo sobre o transdutor minimal com emissões em um monóide cancelativo com $\mathrm{mdc}$ único completa-se com os seguintes resultados: definimos uma congruência $\sim_f$ de uma função entre monóides, que estende o conceito de congruência à direita, ou congruência de Nerode, de uma linguagem; descrevemos um conceito de equivalência $\equiv$ de estados para transdutores sequenciais, que estende o conceito correspondente para autômatos finitos. Com essa ideia, é possível construir um transdutor minimal a partir de uma partição do conjunto de estados de um dado transdutor sequencial. Em se tratando do monóide livre, o conceito da congruência à direita de uma função apareceu no trabalho fundamental de Schützenberger sobre funções sequenciais [Schutzenberger 1977]; os conceitos de estados equivalentes e transdutor quociente estão em [Reutenauer 1990]. Os resultados relevantes nesse contexto são:

**Teorema 3.1** *Sejam $\Sigma$ um alfabeto, e $\mathrm{M}$ um monóide cancelativo com $\mathrm{mdc}$ único. Uma função $f : \Sigma^* \to \mathrm{M}$ é seqüencial se, e somente se, o quociente $\Sigma^*/\sim_f$ é finito.*

**Teorema 3.2** *Seja $\mathcal{T}$ um transdutor sequencial bi-acessível e prefixo que realiza uma função sequencial $f : \Sigma^* \to \mathrm{M}$. Então, $\mathcal{T}/\equiv\ \cong\ \mathcal{T}_f$.*

## Referências


Béal, M. P., Carton, O., Prieur, C., and Sakarovitch, J. (2003). Squaring transducers: an efficient procedure for deciding functionality and sequentiality. *Theoretical Computer Science*, 292(1):45–63.

Benesty, J., Sondhi, M., and Huang, Y., editors (2008). *Springer Handbook of Speech Processing*. Springer-Verlag Berlin Heidelberg.

Choffrut, C. (1979). A generalization of Ginsburg and Rose's characterization of gsm mappings. In *ICALP'79*, volume 71 of *Lecture Notes in Computer Science*, pages 88–103. Springer-Verlag.

Choffrut, C. (2003). Minimizing subsequential transducers: A survey. *Theoretical Computer Science*, 292(1):131–143.

de Souza, R. (2004). Propriedades de algumas classes de relacões racionais. Master's thesis, Instituto de Matemática e Estatística, Universidade de São Paulo.

Eilenberg, S. (1974). *Automata, Languages, and Machines*. Academic Press.

Lombardy, S. and Sakarovitch, J. (2006). Sequential? *Theoretical Computer Science*, 356(1-2):224–244.

Mohri, M. (2000). Minimization algorithms for sequential transducers. *Theoretical Computer Science*, 234(1-2):177–201.

Reutenauer, C. (1990). Subsquential functions: Characterizations, minimization, examples. In *International Meeting of Young Computer Scientists*, volume 464 of *Lecture Notes in Computer Science*, pages 62–79.

Sakarovitch, J. (2009). *Elements of Automata Theory*. Cambridge University Press.

Schutzenberger, M. P. (1977). Sur une variante des fonctions sequentielles. *Theoretical Computer Science*, 4(1):47–57.