



Free (rational) derivation

K. SCHREMPF

*Austrian Academy of Sciences, Acoustics Research Institute
Wohllebengasse 12-14, 1040 Vienna, Austria*

math@versibilitas.at, <https://orcid.org/0000-0001-8509-009X>

Received April 15, 2021
Accepted May 21, 2021

Presented by Manuel Saorín

Abstract: By representing elements in *free fields* (over a *commutative* field and a *finite* alphabet) using Cohn and Reutenauer’s *linear representations*, we provide an algorithmic construction for the (partial) non-commutative (or Hausdorff-) *derivative* and show how it can be applied to the non-commutative version of the Newton iteration to find roots of matrix-valued rational equations.

Key words: Hausdorff derivative, free associative algebra, free field, minimal linear representation, admissible linear system, free fractions, chain rule, Newton iteration.

MSC (2020): primary 16K40, 16S85; secondary 68W30, 46G05.

INTRODUCTION

Working symbolically with matrices requires *non-commuting* variables and thus non-commutative (nc) rational expressions. Although the (algebraic) construction of *free fields*, that is, universal fields of fractions of *free associative algebras*, is available due to Paul M. Cohn since 1970 [8, Chapter 7], its practical application in terms of *free fractions* [29]—building directly on Cohn and Reutenauer’s *linear representations* [10]—in computer algebra systems is only at the very beginning.

The main difficulty for arithmetic—or rather *lexetic* from the non-existing Greek word *λεξητικός* (from *λεξίς* for *word*) as analogon to *αριθμητικός* (from *αριθμός* for *number*)—was the construction of *minimal* linear representations [32], that is, the *normal form* of Cohn and Reutenauer [9].

Here we will show that free fractions also provide a framework for “free” derivation, in particular of nc polynomials. The construction we provide generalizes the univariate (commutative) case we are so much used to, for example

$$f = f(x) = x^3 + 4x^2 + 3x + 5 \quad \text{with} \quad f' = \frac{d}{dx}f(x) = 3x^2 + 8x + 3.$$



The coefficients are from a *commutative* field \mathbb{K} (for example the rational \mathbb{Q} , the real \mathbb{R} , or the complex number field \mathbb{C}), the (non-commuting) variables from a (finite) alphabet \mathcal{X} , for example $\mathcal{X} = \{x, y, z\}$.

For simplicity we focus here (in this motivation) on the *free associative algebra* $R := \mathbb{K}\langle\mathcal{X}\rangle$, aka “algebra of nc polynomials”, and recall the properties of a (partial) *derivation* $\partial_x : R \rightarrow R$ (for a fixed $x \in \mathcal{X}$), namely

- $\partial_x(\alpha) = 0$ for $\alpha \in \mathbb{K}$, and more general, $\partial_x(g) = 0$ for $g \in \mathbb{K}\langle\mathcal{X} \setminus \{x\}\rangle$,
- $\partial_x(x) = 1$, and
- $\partial_x(fg) = \partial_x(f)g + f\partial_x(g)$.

In [27] this is called *Hausdorff derivative*. For a more general (module theoretic) context we refer to [7, Section 2.7] or [1].

Before we continue we should clarify the wording: To avoid confusion we refer to the *linear operator* ∂_x (for some $x \in \mathcal{X}$) as *free partial derivation* (or just *derivation*) and call $\partial_x f \in R$ the (free partial) *derivative* of $f \in R$, sometimes denoted also as f_x or f' depending on the context.

In the commutative, we usually do not distinguish too much between algebraic and analytic concepts. But in the (free) non-commutative setting, analysis is quite subtle [21]. There are even concepts like “matrix convexity” and symbolic procedures to determine (nc) convexity [4]. However, a systematic treatment of the underlying algebraic tools was not available so far. We are going to close this gap in the following.

After a brief description of the setup (in particular that of *linear representations* of elements in free fields) in Section 1, we develop the formalism for *free derivations* in Section 2 with the main result, Theorem 2.8. To be able to state a (partial) “free” chain rule (in Proposition 3.2) we derive a language for the *free composition* (and illustrate how to “reverse” it) in Section 3. And finally, in Section 4, we show how to develop a meta algorithm “nc Newton” to find matrix-valued roots of a non-commutative rational equation.

Notation. The set of the natural numbers is denoted by $\mathbb{N} = \{1, 2, \dots\}$, that including zero by \mathbb{N}_0 . Zero entries in matrices are usually replaced by (lower) dots to emphasize the structure of the non-zero entries unless they result from transformations where there were possibly non-zero entries before. We denote by I_n the identity matrix (of size n) respectively I if the size is clear from the context. By v^\top we denote the transpose of a vector v .

1. GETTING STARTED

We represent elements (in free fields) by *admissible linear systems* (Definition 1.6), which are just a special form of *linear representations* (Definition 1.3) and “general” *admissible systems* [8, Section 7.1]. Rational operations (scalar multiplication, addition, multiplication, inverse) can be easily formulated in terms of linear representations [10, Section 1]. For the formulation on the level of admissible linear systems and the “minimal” inverse we refer to [30, Proposition 1.13] resp. [30, Theorem 4.13].

Let \mathbb{K} be a *commutative* field, $\overline{\mathbb{K}}$ its algebraic closure and $\mathcal{X} = \{x_1, x_2, \dots, x_d\}$ be a *finite* (non-empty) alphabet. $\mathbb{K}\langle\mathcal{X}\rangle$ denotes the *free associative algebra* (or *free \mathbb{K} -algebra*) and $\mathbb{F} = \mathbb{K}\langle\langle\mathcal{X}\rangle\rangle$ its *universal field of fractions* (or “free field”) [6, 10]. An element in $\mathbb{K}\langle\mathcal{X}\rangle$ is called (non-commutative or nc) *polynomial*. In our examples the alphabet is usually $\mathcal{X} = \{x, y, z\}$. Including the algebra of *nc rational series* [2] we have the following chain of inclusions:

$$\mathbb{K} \subsetneq \mathbb{K}\langle\mathcal{X}\rangle \subsetneq \mathbb{K}^{\text{rat}}\langle\langle\mathcal{X}\rangle\rangle \subsetneq \mathbb{K}\langle\langle\mathcal{X}\rangle\rangle =: \mathbb{F}.$$

DEFINITION 1.1. ([8, SECTION 0.1], [10]) Given a matrix $A \in \mathbb{K}\langle\mathcal{X}\rangle^{n \times n}$, the *inner rank* of A is the smallest number $k \in \mathbb{N}$ such that there exists a factorization $A = CD$ with $C \in \mathbb{K}\langle\mathcal{X}\rangle^{n \times k}$ and $D \in \mathbb{K}\langle\mathcal{X}\rangle^{k \times n}$. The matrix A is called *full* if $k = n$, *non-full* otherwise.

THEOREM 1.2. ([8, SPECIAL CASE OF COROLLARY 7.5.14]) *Let \mathcal{X} be an alphabet and \mathbb{K} a commutative field. The free associative algebra $R = \mathbb{K}\langle\mathcal{X}\rangle$ has a universal field of fractions $\mathbb{F} = \mathbb{K}\langle\langle\mathcal{X}\rangle\rangle$ such that every full matrix over R can be inverted over \mathbb{F} .*

Remark. Non-full matrices become singular under a homomorphism into some field [8, Chapter 7]. In general (rings), neither do full matrices need to be invertible, nor do invertible matrices need to be full. An example for the former is the matrix

$$B = \begin{bmatrix} \cdot & z & -y \\ -z & \cdot & x \\ y & -x & \cdot \end{bmatrix}$$

over the *commutative* polynomial ring $\mathbb{K}[x, y, z]$ which is *not* a Sylvester domain [5, Section 4]. An example for the latter are rings without *unbounded generating number* (UGN) [8, Section 7.3].

DEFINITION 1.3. ([9, 10]) Let $f \in \mathbb{F}$. A *linear representation* of f is a triple $\pi_f = (u, A, v)$ with $u^\top, v \in \mathbb{K}^{n \times 1}$, full $A = A_0 \otimes 1 + A_1 \otimes x_1 + \dots + A_d \otimes x_d$ with $A_\ell \in \mathbb{K}^{n \times n}$ for all $\ell \in \{0, 1, \dots, d\}$ and $f = uA^{-1}v$. The *dimension* of π_f is $\dim(u, A, v) = n$. It is called *minimal* if A has the smallest possible dimension among all linear representations of f . The “empty” representation $\pi = (, ,)$ is the minimal one of $0 \in \mathbb{F}$ with $\dim \pi = 0$. Let $f \in \mathbb{F}$ and π be a *minimal* linear representation of f . Then the *rank* of f is defined as $\text{rank } f = \dim \pi$.

DEFINITION 1.4. ([9]) Let $\pi = (u, A, v)$ be a linear representation of $f \in \mathbb{F}$ of dimension n . The families $(s_1, s_2, \dots, s_n) \subseteq \mathbb{F}$ with $s_i = (A^{-1}v)_i$ and $(t_1, t_2, \dots, t_n) \subseteq \mathbb{F}$ with $t_j = (uA^{-1})_j$ are called *left family* and *right family* respectively. $L(\pi) = \text{span}\{s_1, s_2, \dots, s_n\}$ and $R(\pi) = \text{span}\{t_1, t_2, \dots, t_n\}$ denote their linear spans (over \mathbb{K}).

PROPOSITION 1.5. ([9, PROPOSITION 4.7]) A representation $\pi = (u, A, v)$ of an element $f \in \mathbb{F}$ is *minimal* if and only if both, the left family and the right family are \mathbb{K} -linearly independent. In this case, $L(\pi)$ and $R(\pi)$ depend only on f .

Remark. The left family $(A^{-1}v)_i$ (respectively the right family $(uA^{-1})_j$) and the solution vector s of $As = v$ (respectively t of $u = tA$) are used synonymously.

DEFINITION 1.6. ([30]) A linear representation $\mathcal{A} = (u, A, v)$ of $f \in \mathbb{F}$ is called *admissible linear system* (ALS) for f , written also as $As = v$, if $u = e_1 = [1, 0, \dots, 0]$. The element f is then the first component of the (unique) solution vector s . Given a linear representation $\mathcal{A} = (u, A, v)$ of dimension n of $f \in \mathbb{F}$ and invertible matrices $P, Q \in \mathbb{K}^{n \times n}$, the transformed $PAQ = (uQ, PAQ, Pv)$ is again a linear representation (of f). If \mathcal{A} is an ALS, the transformation (P, Q) is called *admissible* if the first row of Q is $e_1 = [1, 0, \dots, 0]$.

2. FREE DERIVATION

Before we define the concrete (partial) *derivation* and (partial) *directional derivation*, we start with a (partial) *formal derivation* on the level of admissible linear systems and show the basic properties with respect to the represented elements, in particular that the (formal) derivation does *not* depend on the ALS in Corollary 2.3.

In other words: Given some letter $x \in \mathcal{X}$ and an *admissible linear system* $\mathcal{A} = (u, A, v)$, there is an *algorithmic* point of view in which the (free) derivation ∂_x defines the ALS $\mathcal{A}' = \partial_x \mathcal{A}$. (Alternatively one can identify x by its index $\ell \in \{1, 2, \dots, d\}$ and write $\mathcal{A}' = \partial_\ell \mathcal{A}$.) Written in a sloppy way, we show that $\partial_x(\mathcal{A} + \mathcal{B}) = \partial_x \mathcal{A} + \partial_x \mathcal{B}$ and $\partial_x(\mathcal{A} \cdot \mathcal{B}) = \partial_x \mathcal{A} \cdot \mathcal{B} + \mathcal{A} \cdot \partial_x \mathcal{B}$, yielding immediately the *algebraic* point of view (summarized in Theorem 2.8) by taking the respective first component of the *unique* solution vectors $f = (s_{\mathcal{A}})_1$ and $g = (s_{\mathcal{B}})_1$.

Remark. The following definition is much more general than usually needed. One gets the “classical” (partial) derivation with respect to some letter $x = x_\ell \in \mathcal{X}$ (with $\ell \in \{1, 2, \dots, d\}$) for $k = 0$ resp. (the empty word) $a = 1 \in \mathcal{X}^*$.

DEFINITION 2.1. Let $\mathcal{A} = (u, A, v)$ be an *admissible linear system* of dimension $n \geq 1$ for some element in the free field $\mathbb{F} = \mathbb{K}\langle\langle \mathcal{X} \rangle\rangle$ and $\ell \neq k \in \{0, 1, \dots, d\}$. The ALS

$$\partial_{\ell|k} \mathcal{A} = \partial_{\ell|k}(u, A, v) = \left([u \quad \cdot], \begin{bmatrix} A & A_\ell \otimes x_k \\ \cdot & A \end{bmatrix}, \begin{bmatrix} \cdot \\ v \end{bmatrix} \right)$$

(of dimension $2n$) is called (partial) *formal derivative* of \mathcal{A} , (with respect to $x_\ell, x_k \in \{1\} \cup \mathcal{X}$). For $x, a \in \{1\} \cup \mathcal{X} = \{1, x_1, x_2, \dots, x_d\}$ with $x \neq a$ we write also $\partial_{x|a} \mathcal{A}$, having the indices $\ell \neq k \in \{0, 1, \dots, d\}$ of x resp. a in mind.

LEMMA 2.2. Let $\mathcal{A}_f = (u_f, A_f, v_f)$ and $\mathcal{A}_g = (u_g, A_g, v_g)$ be *admissible linear systems* of dimension $\dim \mathcal{A}_f \geq 1$ resp. $\dim \mathcal{A}_g \geq 1$. Fix $x, a \in \{1\} \cup \mathcal{X}$ such that $x \neq a$. Then $\partial_{x|a}(\mathcal{A}_f + \mathcal{A}_g) = \partial_{x|a} \mathcal{A}_f + \partial_{x|a} \mathcal{A}_g$.

Proof. Let $\ell \neq k \in \{0, 1, \dots, d\}$ be the indices of x resp. a . We write A_f^ℓ for the coefficient matrix A_ℓ of \mathcal{A}_f resp. A_g^ℓ for A_ℓ of \mathcal{A}_g . Taking the sum from [30, Proposition 1.13] we have

$$\begin{aligned} \partial_{x|a}(\mathcal{A}_f + \mathcal{A}_g) &= \left([u_f \quad \cdot], \begin{bmatrix} A_f & -A_f u_f^\top u_g \\ \cdot & A_g \end{bmatrix}, \begin{bmatrix} v_f \\ v_g \end{bmatrix} \right) \\ &= \left([u_f \quad \cdot \quad \cdot \quad \cdot], \begin{bmatrix} A_f & -A_f u_f^\top u_g & A_f^\ell \otimes a & -A_f^\ell u_f^\top u_g \otimes a \\ \cdot & A_g & \cdot & A_g^\ell \otimes a \\ \cdot & \cdot & A_f & -A_f u_f^\top u_g \\ \cdot & \cdot & \cdot & A_g \end{bmatrix}, \begin{bmatrix} \cdot \\ \cdot \\ v_f \\ v_g \end{bmatrix} \right) \end{aligned}$$

$$\begin{aligned}
&= \left([u_f \ \cdot \ \cdot \ \cdot], \begin{bmatrix} A_f & A_f^\ell \otimes a & -A_f u_f^\top u_g & -A_f^\ell u_f^\top u_g \otimes a \\ \cdot & A_f & \cdot & -A_f u_f^\top u_g \\ \cdot & \cdot & A_g & A_g^\ell \otimes a \\ \cdot & \cdot & \cdot & A_g \end{bmatrix}, \begin{bmatrix} \cdot \\ v_f \\ \cdot \\ v_g \end{bmatrix} \right) \\
&= \left([u_f \ \cdot \ \cdot \ \cdot], \begin{bmatrix} A_f & A_f^\ell \otimes a & -A_f u_f^\top u_g & 0 \\ \cdot & A_f & \cdot & 0 \\ \cdot & \cdot & A_g & A_g^\ell \otimes a \\ \cdot & \cdot & \cdot & A_g \end{bmatrix}, \begin{bmatrix} \cdot \\ v_f \\ \cdot \\ v_g \end{bmatrix} \right) \\
&= \left([u_f \ \cdot], \begin{bmatrix} A_f & A_f^\ell \otimes a \\ \cdot & A_f \end{bmatrix}, \begin{bmatrix} \cdot \\ v_f \end{bmatrix} \right) + \left([u_g \ \cdot], \begin{bmatrix} A_g & A_g^\ell \otimes a \\ \cdot & A_g \end{bmatrix}, \begin{bmatrix} \cdot \\ v_g \end{bmatrix} \right) \\
&= \partial_{x|a} \mathcal{A}_f + \partial_{x|a} \mathcal{A}_g.
\end{aligned}$$

The two main steps are swapping block rows 2 and 3 and block columns 2 and 3, and eliminating the *single* non-zero (first) column in $-A_f^\ell u_f^\top u_g \otimes a$ and $-A_f u_f^\top u_g$ (in block column 4) using the first column in block column 2. ■

COROLLARY 2.3. *Let $f, g \in \mathbb{F}$ be given by the admissible linear systems $\mathcal{A}_f = (u_f, A_f, v_f)$ and $\mathcal{A}_g = (u_g, A_g, v_g)$ of dimensions $n_f, n_g \geq 1$ respectively. Fix $x, a \in \{1\} \cup \mathcal{X}$ such that $x \neq a$. Then $f = g$ implies that $\partial_{x|a} \mathcal{A}_f - \partial_{x|a} \mathcal{A}_g$ is an ALS for $0 \in \mathbb{F}$.*

DEFINITION 2.4. Let $f \in \mathbb{F}$ be given by the ALS $\mathcal{A} = (u, A, v)$ and fix $x, a \in \{1\} \cup \mathcal{X} = \{1, x_1, x_2, \dots, x_d\}$ such that $x \neq a$. Denote by $\partial_{x|a} f$ the element defined by the ALS $\partial_{x|a} \mathcal{A}$. The map $\partial_{x|a} : \mathbb{F} \rightarrow \mathbb{F}$, $f \mapsto \partial_{x|a} f$ is called (partial) *formal derivation*, the element $\partial_{x|a} f$ (partial) *formal derivative* of f .

COROLLARY 2.5. *For each $x, a \in \{1\} \cup \mathcal{X}$ with $x \neq a$, the formal derivation $\partial_{x|a} : \mathbb{F} \rightarrow \mathbb{F}$ is a linear map.*

Now we are almost done. Before we show the *product rule* in the following Lemma 2.6, we have a look into the left family of the ALS of the (formal) derivative of a polynomial. Let $p = x^3 \in \mathbb{F}$ (and $a = 1$). A (minimal) ALS for $\partial_{x|1} p$ is given by

$$\begin{bmatrix} 1 & -x & \cdot & \cdot & 0 & -1 & \cdot & \cdot \\ \cdot & 1 & -x & \cdot & \cdot & 0 & -1 & \cdot \\ \cdot & \cdot & 1 & -x & \cdot & \cdot & 0 & -1 \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & 1 & -x & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & -x & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & -x \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix} s = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}, \quad s = \begin{bmatrix} 3x^2 \\ 2x \\ 1 \\ 0 \\ x^3 \\ x^2 \\ x \\ 1 \end{bmatrix}.$$

Notice that the first four entries in the left family of $\partial_{x|1}\mathcal{A}$ are $s_i = \partial_{x|1}s_{i+4}$.

LEMMA 2.6. (PRODUCT RULE) *Let $f, g \in \mathbb{F}$ be given by the admissible linear systems $\mathcal{A}_f = (u_f, A_f, v_f)$ and $\mathcal{A}_g = (u_g, A_g, v_g)$ of dimension $n_f, n_g \geq 1$ respectively. Fix $x \in \mathcal{X}$ and $a \in \{1\} \cup \mathcal{X} \setminus \{x\}$. Then*

$$\partial_{x|a}(fg) = \partial_{x|a}f g + f \partial_{x|a}g.$$

Proof. Let $\ell \neq k \in \{0, 1, \dots, d\}$ be the indices of x resp. a . We write A_f^ℓ for the coefficient matrix A_ℓ of A_f resp. A_g^ℓ for A_ℓ of A_g . We take the sum and the product from [30, Proposition 1.13] and start with the ALS from the right hand side,

$$\begin{bmatrix} A_f & A_f^\ell \otimes a & 0 & -A_f u_f^\top u_g & \cdot & \cdot \\ \cdot & A_f & -v_f u_g & \cdot & \cdot & \cdot \\ \cdot & \cdot & A_g & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & A_f & -v_f u_g & \cdot \\ \cdot & \cdot & \cdot & \cdot & A_g & A_g^\ell \otimes a \\ \cdot & \cdot & \cdot & \cdot & \cdot & A_g \end{bmatrix} s = \begin{bmatrix} \cdot \\ \cdot \\ v_g \\ \cdot \\ \cdot \\ v_g \end{bmatrix},$$

subtract block row 6 from block row 3, add block column 3 to block column 6 and remove block row/column 3 to get the ALS

$$\begin{bmatrix} A_f & A_f^\ell \otimes a & -A_f u_f^\top u_g & \cdot & \cdot \\ \cdot & A_f & \cdot & \cdot & -v_f u_g \\ \cdot & \cdot & A_f & -v_f u_g & \cdot \\ \cdot & \cdot & \cdot & A_g & A_g^\ell \otimes a \\ \cdot & \cdot & \cdot & \cdot & A_g \end{bmatrix} s = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ v_g \end{bmatrix}.$$

Now we can add block row 3 to block row 1 and eliminate the remaining columns in block (1, 3) by the columns $\{2, 3, \dots, n_f\}$ from block (1, 1), remove block row/column 3 to get the ALS

$$\begin{bmatrix} A_f & A_f^\ell \otimes a & -v_f u_g & \cdot \\ \cdot & A_f & \cdot & -v_f u_g \\ \cdot & \cdot & A_g & A_g^\ell \otimes a \\ \cdot & \cdot & \cdot & A_g \end{bmatrix} s = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ v_g \end{bmatrix}.$$

Swapping block rows 2 and 3 and block columns 2 and 3 yields the ALS

$$\begin{bmatrix} A_f & -v_f u_g & A_f^\ell \otimes a & 0 \\ \cdot & A_g & \cdot & A_g^\ell \otimes a \\ \cdot & \cdot & A_f & -v_f u_g \\ \cdot & \cdot & \cdot & A_g \end{bmatrix} s = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ v_g \end{bmatrix}$$

of the left hand side $\partial_{x|0}(fg)$. Notice the upper right zero in the system matrix which is because of $x \neq 1$. ■

DEFINITION 2.7. Let $f \in \mathbb{F}$, $x \in \mathcal{X}$ and $a \in \mathcal{X} \setminus \{x\}$. The element $\partial_x f := \partial_{x|1} f$ is called *partial derivative* of f . The element $\partial_{x|a} f$ is called (partial) *directional derivative* of f (with respect to a).

THEOREM 2.8. (FREE DERIVATION) Let $x \in \mathcal{X}$. Then the (partial) free derivation $\partial_x : \mathbb{F} \rightarrow \mathbb{F} = \mathbb{K}(\langle \mathcal{X} \rangle)$ is the unique map with the properties

- $\partial_x h = 0$ for all $h \in \mathbb{K}(\langle \mathcal{X} \setminus \{x\} \rangle)$,
- $\partial_x x = 1$, and
- $\partial_x(fg) = \partial_x f g + f \partial_x g$ for all $f, g \in \mathbb{F} = \mathbb{K}(\langle \mathcal{X} \rangle)$.

Proof. Let h be given by the ALS $\mathcal{A} = (u, A, v)$ and let $\ell \in \{1, 2, \dots, d\}$ such that $x = x_\ell$. We just need to recall the ALS for $\partial_x h$,

$$\begin{bmatrix} A & A_\ell \otimes 1 \\ \cdot & A \end{bmatrix} \begin{bmatrix} s' \\ s'' \end{bmatrix} = \begin{bmatrix} \cdot \\ v \end{bmatrix}$$

and observe that $A_\ell = 0$ and thus $As' = 0$, in particular the first component of s' . Therefore $\partial_x h = 0$. For $\partial_x x = 1$ we need to minimize

$$\begin{bmatrix} 1 & -x & \cdot & -1 \\ \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & 1 & -x \\ \cdot & \cdot & \cdot & 1 \end{bmatrix} s = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}.$$

And the product rule $\partial_x(fg) = \partial_x f g + f \partial_x g$ is due to Lemma 2.6. For the uniqueness we assume that there exists another $\partial'_x : \mathbb{F} \rightarrow \mathbb{F}$ with the same properties. From the product rule we obtain

$$\partial_x(xf) = f + x \partial_x f = f + x \partial'_x f = \partial'_x(xf),$$

that is, $x(\partial_x f - \partial'_x f) = 0$ for all $f \in \mathbb{F}$, thus $\partial_x = \partial'_x$. ■

COROLLARY 2.9. (HAUSDORFF DERIVATION [27]) *Let $x \in \mathcal{X}$. Then $\partial_x \kappa = 0$ for all $\kappa \in \mathbb{K}$, $\partial_x y = 0$ for all $y \in \mathcal{X} \setminus \{x\}$, $\partial_x x = 1$, and $\partial_x(fg) = \partial_x f g + f \partial_x g$ for all $f, g \in \mathbb{K}\langle \mathcal{X} \rangle$.*

Remark. More general [8, Theorem 7.5.17]: “Any derivation of a Sylvester domain extends to a derivation of its universal field of fractions.” Recall however that the *cyclic derivative* is *not* (from) a derivation [27, Section 1]. For a discussion of cyclic derivatives of nc algebraic power series we refer to [24].

PROPOSITION 2.10. *Let $f \in \mathbb{F}$, $x, y \in \mathcal{X}$ and $a, b \in \mathcal{X} \setminus \{x, y\}$ with $a = b$ if and only if $x = y$. Then $\partial_{x|a}(\partial_{y|b}f) = \partial_{y|b}(\partial_{x|a}f)$, that is, the (partial) derivations $\partial_{x|a}$ and $\partial_{y|b}$ commute.*

Proof. Let $l, k \in \{1, 2, \dots, d\}$ the indices of x resp. y . There is nothing to show for the trivial case $x = y$, thus we can assume $l \neq k$. Let f be given by the admissible linear system $\mathcal{A} = (u, A, v)$. Then the “left” ALS $\partial_{x|a}(\partial_{y|b}\mathcal{A})$ is

$$\begin{bmatrix} A & A_k \otimes b & A_l \otimes a & 0 \\ \cdot & A & 0 & A_l \otimes a \\ \cdot & \cdot & A & A_k \otimes b \\ \cdot & \cdot & \cdot & A \end{bmatrix} s = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ v \end{bmatrix}.$$

Swapping block rows/columns 2 and 3 yields the “right” ALS $\partial_{y|b}(\partial_{x|a}\mathcal{A})$:

$$\begin{bmatrix} A & A_l \otimes a & A_k \otimes b & 0 \\ \cdot & A & 0 & A_k \otimes b \\ \cdot & \cdot & A & A_l \otimes a \\ \cdot & \cdot & \cdot & A \end{bmatrix} s = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ v \end{bmatrix}.$$

■

Since there is no danger of ambiguity, we can define the (free) “higher” derivative of $f \in \mathbb{F}$ as $\partial_w f$ for each word w in the *free monoid* \mathcal{X}^* with the “trivial” derivative $\partial_{(1)} f = f$. Let $w \in \mathcal{X}^*$ and $\sigma(w)$ denote any permutation of the letters of w . Then $\partial_w f = \partial_{\sigma(w)} f$.

The proof for nc formal power series in [23, Proposition 1.8] is based on words (monomials), that is, $\partial_x(\partial_y w) = \partial_y(\partial_x w) \in \mathbb{K}\langle\mathcal{X}\rangle$. Recall that one gets the (nc) *rational* series by intersecting the (nc) series and the *free field* [26, Section 9]:

$$\mathbb{K}\langle\mathcal{X}\rangle \cap \mathbb{K}(\langle\mathcal{X}\rangle) = \mathbb{K}^{\text{rat}}\langle\mathcal{X}\rangle.$$

Overall, (free) nc derivation does not appear that often in the literature. And when there is some discussion it is (almost) always connected with “not simple” [27, Section 1], “complicated” [16, Section 14.3], etc. This is however *not* due to the Hausdorff derivation but to the use of (finite) formal series as representation (for nc polynomials). Using *linear representations* in the sense of Cohn and Reutenauer [9] for elements in the free field \mathbb{F} can even reveal additional structure, as indicated in Example 2.11 (below). For the somewhat more “complicated” example

$$\begin{aligned} \tilde{p} = & 3cyxb + 3xbyxb + 2cyxax + cybxb - cyaxb - 2xbyxax + 4xbybxb \\ & - 3xbyaxb + 3xaxyxb - 3bxybxb + 6axbyxb + 2xaxyxax + xaxybxb \\ & - xaxyaxb - 2bxyxax - bxybxb + bxyaxb + 5axbybxb - 4axbyaxb \end{aligned}$$

from [3, Section 8.2] we refer to [31, Example 3.7].

EXAMPLE 2.11. Let $p = xyzx$. A (minimal) polynomial ALS for p is

$$\begin{bmatrix} 1 & -x & \cdot & \cdot & \cdot \\ \cdot & 1 & -y & \cdot & \cdot \\ \cdot & \cdot & 1 & -z & \cdot \\ \cdot & \cdot & \cdot & 1 & -x \\ \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix} s = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}.$$

Then $\partial_x p = xyz + yzx$ admits a factorization into matrices [31, Section 3]:

$$\partial_x p = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} y & \cdot \\ \cdot & z \end{bmatrix} \begin{bmatrix} z \\ x \end{bmatrix}.$$

A *minimal ALS* for $\partial_x p$ is

$$\begin{bmatrix} 1 & -x & -y & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & -y & 0 & \cdot \\ \cdot & \cdot & 1 & 0 & -z & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & -z \\ \cdot & \cdot & \cdot & \cdot & 1 & -x \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix} s = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}.$$

EXAMPLE 2.12. We will use the *directional derivative* later in Section 4 for the (nc) Newton iteration. For $q = x^2$ we get the “Sylvester equation” $\partial_x|_a q = xa + ax$ which is *linear* in a .

In (the next) Section 3 we have a look on the “free” chain rule which will turn out to be very elegant. We avoid the term “function” here since one needs to be careful with respect to evaluation (domain of definition), e.g. $f = (xy - yx)^{-1}$ is not defined for diagonal matrices. For the efficient evaluation of polynomials (by matrices) one can use *Horner Systems* [31]. A generalization to elements in free fields is considered in future work. If there is a “compositional structure” available (in admissible linear systems) it could be used to further optimize evaluation.

Remark. For details on minimization (of linear representations) we refer to [32]. Notice in particular that the construction (of the formal derivative) in Definition 2.1 *preserves* refined pivot blocks. Therefore, if \mathcal{A} is refined, *linear* (algebraic) techniques suffice for minimization of $\partial_x \mathcal{A}$.

Last but not least, given the alphabet $\mathcal{X} = \{x_1, x_2, \dots, x_d\}$, we can define the “free” (canonical) *gradient*

$$\nabla f = [\partial_1 f, \partial_2 f, \dots, \partial_d f]^\top = [\partial_{x_1} f, \partial_{x_2} f, \dots, \partial_{x_d} f]^\top \in \mathbb{F}^d$$

for some $f \in \mathbb{K}\langle\langle \mathcal{X} \rangle\rangle$. *Cyclic* gradients are discussed in [34].

And for a “vector valued” element $\mathbf{f} = (f_1, f_2, \dots, f_d)$ we can define the *Jacobian matrix* $J(\mathbf{f}) = (\partial_j f_i)_{i,j=1}^d = (\partial_{x_j} f_i)_{i,j=1}^d$. For a discussion of non-commutative Jacobian matrices on the level of nc formal power series we refer to [25].

3. FREE COMPOSITION

To be able to formulate a (partial) “free” *chain rule* in an elegant way, we need a suitable notation. It will turn out that Cohn’s *admissible systems* [8, Section 7.1] provide the perfect framework for the “expansion” of letters by elements from another free field.

First we recall the “classical” (analytical) chain rule: Let X, Y and Z be (open) sets, $f = f(x)$ differentiable on X , $g = g(y)$ differentiable on Y and $h = h(x) = g(f(x))$. Then $\frac{d}{dx}h = \frac{d}{df}g \frac{d}{dx}f$ resp. $h'(x) = g'(f(x))f'(x)$.

$$\begin{array}{ccccc} X & \xrightarrow{f} & Y & \xrightarrow{g} & Z \\ & & \searrow & \nearrow & \\ & & & & h \end{array}$$

Notation. For a fixed $d \in \mathbb{N}$ let $\mathcal{X} = \{x_1, x_2, \dots, x_d\}$, $\mathcal{Y} = \{y_1, y_2, \dots, y_d\}$ and $\mathcal{Y}' = \{y'_1, y'_2, \dots, y'_d\}$ be *pairwise disjoint* alphabets, that is,

$$\mathcal{X} \cap \mathcal{Y} = \mathcal{X} \cap \mathcal{Y}' = \mathcal{Y} \cap \mathcal{Y}' = \emptyset.$$

By $\mathbb{F}_{\mathcal{Z}}$ we denote the free field $\mathbb{K}(\langle \mathcal{Z} \rangle)$. Let $A \in \mathbb{K}(\mathcal{Y})^{n \times n}$ be a *linear full* matrix and $\mathbf{f} = (f_1, f_2, \dots, f_d)$ a d -tuple of elements $f_i \in \mathbb{F}_{\mathcal{X}}$. By $A \circ \mathbf{f}$ we denote the (not necessarily full) $n \times n$ matrix over $\mathbb{F}_{\mathcal{X}}$ where each letter $y_i \in \mathcal{Y}$ is replaced by the corresponding $f_i \in \mathbb{F}_{\mathcal{X}}$, that is,

$$A \circ \mathbf{f} = A_0 \otimes 1 + A_1 \otimes f_1 + \dots + A_d \otimes f_d \in \mathbb{F}_{\mathcal{X}}^{n \times n}.$$

We write $A \circ \mathcal{A}_{\mathbf{f}} = A \circ (\mathcal{A}_{f_1}, \mathcal{A}_{f_2}, \dots, \mathcal{A}_{f_d})$ for a *linearized* version induced by the d -tuple of admissible linear systems $\mathcal{A}_{f_i} = (u_{f_i}, A_{f_i}, v_{f_i})$.

Now let $g \in \mathbb{F}_{\mathcal{Y}}$ be given by the ALS $\mathcal{A}_g = (u_g, A_g, v_g)$ and $\mathbf{f} = (f_1, f_2, \dots, f_d) \in \mathbb{F}_{\mathcal{X}}^d$ such that $A_g \circ \mathbf{f}$ is *full*. Then (the *unique* element) $h = g \circ \mathbf{f} \in \mathbb{F}_{\mathcal{X}}$ is defined by the *admissible system* $\tilde{\mathcal{A}}_h = (u_g, A_g \circ \mathbf{f}, v_g)$ and we write

$$\mathcal{A}_h = (u_h, A_h, v_h) = (u_g \circ \mathcal{A}_{\mathbf{f}}, A_g \circ \mathcal{A}_{\mathbf{f}}, v_g \circ \mathcal{A}_{\mathbf{f}}) =: \mathcal{A}_g \circ \mathcal{A}_{\mathbf{f}}$$

for a *linearized* version using “linearization by enlargement” [8, Section 5.8]. (For details we refer to the proof of Proposition 3.2 below.) Fixing some $x \in \mathcal{X}$, we get the (partial) derivative $\partial_x h$ of h via the derivative $\partial_x \mathcal{A}_h$

$$\begin{bmatrix} A_h & A_h^x \otimes 1 \\ \cdot & A_h \end{bmatrix} s = \begin{bmatrix} \cdot \\ v_h \end{bmatrix}.$$

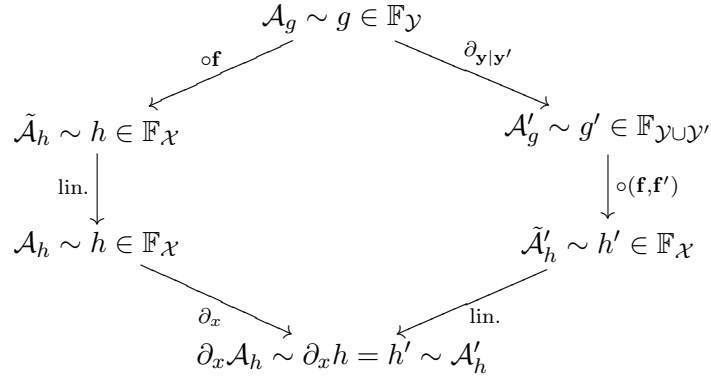


Figure 1: Let $\mathbf{f} = (f_1, \dots, f_d)$ with $f_i \in \mathbb{F}_{\mathcal{X}} = \mathbb{K}\langle\langle\mathcal{X}\rangle\rangle$ given by the d -tuple of admissible linear systems $\mathcal{A}_{\mathbf{f}} = (\mathcal{A}_1, \dots, \mathcal{A}_d)$ and $g \in \mathbb{F}_{\mathcal{Y}}$ given by $\mathcal{A}_g = (u_g, A_g, v_g)$ such that the system matrix A_g remains *full* when we replace each letter $y_i \in \mathcal{Y}$ by the respective element f_i , written as $A_g \circ \mathbf{f}$. Then $h = g \circ \mathbf{f} \in \mathbb{F}_{\mathcal{X}}$ is defined by the *admissible system* $\tilde{\mathcal{A}}_h = (u_g, A_g \circ \mathbf{f}, v_g)$ which we *linearize* to obtain an ALS \mathcal{A}_h for h before we apply the (partial) derivation ∂_x (left path). On the other hand, we can apply the (directional) derivation $\partial_{\mathcal{Y}|\mathcal{Y}'}$ by going over to the free field $\mathbb{F}_{\mathcal{Y} \cup \mathcal{Y}'} = \mathbb{K}\langle\langle\mathcal{Y} \cup \mathcal{Y}'\rangle\rangle$ with an extended alphabet (with “placeholders” y'_i), yielding $g' = \partial_{y_1|y'_1}g + \dots + \partial_{y_d|y'_d}g$ given by some ALS $\mathcal{A}'_g = (u'_g, A'_g, v'_g)$. Then $h' = g' \circ (\mathbf{f}, \mathbf{f}') \in \mathbb{F}_{\mathcal{X}}$ is defined by the *admissible system* $\tilde{\mathcal{A}}'_h = (u'_g, A'_g \circ (\mathbf{f}, \mathbf{f}'), v'_g)$, where also each letter $y'_i \in \mathcal{Y}'$ is replaced by the respective element $f'_i = \partial_x f_i \in \mathbb{F}_{\mathcal{X}}$. After linearization we get an ALS \mathcal{A}'_h such that $\partial_x \mathcal{A}_h - \mathcal{A}'_h = 0$, that is, $\partial_x h = h'$ (right path).

To give a meaning to the right hand side of $\partial_x h = \partial_x(g \circ \mathbf{f})$, we introduce the “total” (directional) derivative

$$g' := \partial_{\mathcal{Y}|\mathcal{Y}'}g = \partial_{y_1|y'_1}g + \partial_{y_2|y'_2}g + \dots + \partial_{y_d|y'_d}g \in \mathbb{F}_{\mathcal{Y} \cup \mathcal{Y}'}$$

given by the ALS

$$\mathcal{A}'_g := \partial_{\mathcal{Y}|\mathcal{Y}'}\mathcal{A}_g = \left([u_g \quad \cdot], \begin{bmatrix} A_g & \sum_{i=1}^d A_g^{(i)} \otimes y'_i \\ \cdot & A_g \end{bmatrix}, \begin{bmatrix} \cdot \\ v_g \end{bmatrix} \right)$$

with letters $y'_i \in \mathcal{Y}'$. Using a similar notation for the derivatives

$$\mathbf{f}' = (f'_1, f'_2, \dots, f'_d) := (\partial_x f_1, \partial_x f_2, \dots, \partial_x f_d) = \partial_x \mathbf{f},$$

we can write

$$h' := \partial_x(g \circ \mathbf{f}) = g' \circ (\mathbf{f}, \mathbf{f}') = \partial_{\mathcal{Y}|\mathcal{Y}'}g \circ (\mathbf{f}, \mathbf{f}') \quad (3.1)$$

given by the *admissible system*

$$\mathcal{A}'_g \circ (\mathbf{f}, \mathbf{f}') = \partial_{\mathcal{Y}|\mathcal{Y}'} \mathcal{A}_g \circ (\mathbf{f}, \mathbf{f}').$$

After an illustration in the following example, we show in Proposition 3.2 that indeed

$$\partial_x h = h' = \partial_x(g \circ \mathbf{f}) = \partial_{\mathcal{Y}|\partial_x} g \circ \mathbf{f}$$

using an abbreviation for the right hand side of (3.1). For an overview see Figure 1.

Remark. Cohn writes an *admissible system* $\mathcal{A} = (u, A, v)$ as “block” $[A, v]$ with not necessarily scalar column v [8, Section 7.1]. A ring homomorphism which preserves fullness of matrices is called *honest* [8, Section 5.4].

Remark. It is *crucial* that $A_g \circ \mathbf{f}$ is *full* to be able to define the composition. On a purely algebraic level this is sufficient to define the (partial) chain rule. For a d -tuple $\mathbf{g} = (g_1, g_2, \dots, g_d)$ given by admissible linear systems $\mathcal{A}_{\mathbf{g}} = (\mathcal{A}_{g_1}, \mathcal{A}_{g_2}, \dots, \mathcal{A}_{g_d})$ with $\mathcal{A}_{g_i} = (u_{g_i}, A_{g_i}, v_{g_i})$ we need $A_{g_i} \circ \mathbf{f}$ *full* for all $i \in \{1, 2, \dots, d\}$.

EXAMPLE 3.1. Here we take $\mathcal{X} = \{x, y\}$, $\mathcal{Y} = \{f, p\}$ and abuse notation. Let $f = (x^{-1} + y)^{-1} \in \mathbb{F}_{\mathcal{X}}$, $p = xy \in \mathbb{F}_{\mathcal{X}}$ and $g = pfp \in \mathbb{F}_{\mathcal{Y}}$. Then $h = g \circ (f, p) \in \mathbb{F}_{\mathcal{X}}$ is given by the (minimal) ALS

$$\begin{bmatrix} 1 & -x & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & -y & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & -x & \cdot & \cdot \\ \cdot & \cdot & y & 1 & -x & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & -y \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix} s = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix},$$

and

$$\partial_x h = \partial_x(pfp) = \partial_x p f p + p \partial_x f p + p f \partial_x p$$

by

$$\left[\begin{array}{cccccc|cccc}
 1 & -x & \cdot & \cdot & \cdot & \cdot & \cdot & -1 & \cdot & \cdot & \cdot & \cdot \\
 \cdot & 1 & -y & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 \cdot & \cdot & 1 & -x & \cdot & \cdot & \cdot & \cdot & \cdot & -1 & \cdot & \cdot \\
 \cdot & \cdot & y & 1 & -x & \cdot & \cdot & \cdot & \cdot & \cdot & -1 & \cdot \\
 \cdot & \cdot & \cdot & \cdot & 1 & -y & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 \hline
 & & & & & & 1 & -x & \cdot & \cdot & \cdot & \cdot \\
 & & & & & & \cdot & 1 & -y & \cdot & \cdot & \cdot \\
 & & & & & & \cdot & \cdot & 1 & -x & \cdot & \cdot \\
 & & & & & & \cdot & \cdot & y & 1 & -x & \cdot \\
 & & & & & & \cdot & \cdot & \cdot & \cdot & 1 & -y \\
 & & & & & & \cdot & \cdot & \cdot & \cdot & \cdot & 1
 \end{array} \right] s = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}. \quad (3.2)$$

On the other hand, $\partial_x(\partial_f g + \partial_p g) = \partial_x(\partial_f g) + \partial_x(\partial_p g)$ is given by (the admissible system)

$$\left[\begin{array}{cccccc|cccc}
 1 & -p & \cdot & \cdot & \cdot & -\partial_x p & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 \cdot & 1 & -f & \cdot & \cdot & \cdot & -\partial_x f & \cdot & \cdot & \cdot & \cdot & \cdot \\
 \cdot & \cdot & 1 & -p & \cdot & \cdot & \cdot & -\partial_x p & \cdot & \cdot & \cdot & \cdot \\
 \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 \cdot & \cdot & \cdot & \cdot & 1 & -p & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 \cdot & \cdot & \cdot & \cdot & \cdot & 1 & -f & \cdot & \cdot & \cdot & \cdot & \cdot \\
 \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & -p & \cdot & \cdot & \cdot & \cdot \\
 \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot
 \end{array} \right] s = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}.$$

The summands $\partial_x p f p$ and $p f \partial_x p$ are easy to read off in the ALS (3.2). (Alternatively one could add row 8 to row 1 resp. row 11 to row 4.) To read off $p \partial_x f p = -p(x^{-1} + y)^{-2} p$, we just need to recall the (minimal) ALS

$$\left[\begin{array}{cccc}
 1 & -x & \cdot & -1 \\
 y & 1 & \cdot & \cdot \\
 & & 1 & -x \\
 & & y & 1
 \end{array} \right] s = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}.$$

In other words and with $g \in \mathbb{F}_y$ given by the ALS $\mathcal{A}_g = (u_g, A_g, v_g)$ of dimension n : $\partial_x h$ is given by the *admissible system* $\partial_y|_{\partial_x} \mathcal{A}_g$,

$$\left[\begin{array}{cc}
 A_g & \partial_x A_g \\
 \cdot & A_g
 \end{array} \right] s = \begin{bmatrix} \cdot \\ v \end{bmatrix}$$

of dimension $2n$. Notice that $\partial_x A_g$ is understood here in a purely symbolic way, that is, over $\mathbb{F}_{\mathcal{Y} \cup \mathcal{Y}'}$ with additional letters “ $\partial_x y$ ” in \mathcal{Y}' for each $y \in \mathcal{Y}$. In this sense $\partial_{\mathcal{Y}|\partial_x} \mathcal{A}_g$ is actually *linear*.

PROPOSITION 3.2. (FREE CHAIN RULE) *Let $\mathcal{X} = \{x_1, x_2, \dots, x_d\}$, $\mathcal{Y} = \{y_1, y_2, \dots, y_d\}$ and $\mathcal{Y}' = \{y'_1, y'_2, \dots, y'_d\}$ be pairwise disjoint alphabets and fix $x \in \mathcal{X}$. For $g \in \mathbb{K}\langle\langle \mathcal{Y} \rangle\rangle$ given by the ALS $\mathcal{A}_g = (u_g, A_g, v_g)$ and the d -tuple $\mathbf{f} = (f_1, f_2, \dots, f_d) \in \mathbb{K}\langle\langle \mathcal{X} \rangle\rangle^d$ such that $A_g \circ \mathbf{f}$ is full. Denote by \mathbf{f}' the d -tuple $(\partial_x f_1, \partial_x f_2, \dots, \partial_x f_d)$. Then*

$$\partial_x(g \circ \mathbf{f}) = \partial_{\mathcal{Y}|\mathcal{Y}'} g \circ (\mathbf{f}, \mathbf{f}') =: \partial_{\mathcal{Y}|\partial_x} g \circ \mathbf{f}.$$

Proof. For $i \in \{1, 2, \dots, d\}$ let $f_i \in \mathbb{F}_{\mathcal{X}}$ be given by the admissible linear systems $\mathcal{A}_{f_i} = (u_{f_i}, A_{f_i}, v_{f_i})$ respectively. In the following we assume —without loss of generality— $d = 3$, and decompose $A_g \circ \mathbf{f}$ into $\begin{bmatrix} A_{11} & A_{12} \\ A_{12} & a \end{bmatrix}$ of size n with $a = \alpha_0 + \alpha_1 f_1 + \alpha_2 f_2 + \alpha_3 f_3$. A generalization of the “linearization by enlargement” [8, Section 5.8] is then to start with the *full* matrix $A_g \circ \mathbf{f} \oplus A_{f_1} \oplus \dots \oplus A_{f_d}$, add $u_{f_i} A_{f_i}^{-1}$ from the corresponding block row to row n , and $-\alpha_i A_{f_i}^{-1} v_{f_i}$ from the corresponding block column to column n in the upper left block (these transformations preserve fullness):

$$\begin{bmatrix} A_{11} & A_{12} & \cdot & \cdot & \cdot \\ A_{21} & \alpha_0 & u_{f_1} & u_{f_2} & u_{f_3} \\ \cdot & -\alpha_1 v_{f_1} & A_{f_1} & \cdot & \cdot \\ \cdot & -\alpha_2 v_{f_2} & \cdot & A_{f_2} & \cdot \\ \cdot & -\alpha_3 v_{f_3} & \cdot & \cdot & A_{f_3} \end{bmatrix}.$$

A *partially* linearized system matrix for $\partial_x(\mathcal{A}_g \circ \mathbf{f})$ is

$$\left[\begin{array}{ccccc|ccccc} A_{11} & A_{12} & \cdot & \cdot & \cdot & * & * & \cdot & \cdot & \cdot \\ A_{21} & \alpha_0 & u_{f_1} & u_{f_2} & u_{f_3} & * & 0 & 0 & 0 & 0 \\ \cdot & -\alpha_1 v_{f_1} & A_{f_1} & \cdot & \cdot & \cdot & 0 & A_{f_1}^x \otimes 1 & \cdot & \cdot \\ \cdot & -\alpha_2 v_{f_2} & \cdot & A_{f_2} & \cdot & \cdot & 0 & \cdot & A_{f_2}^x \otimes 1 & \cdot \\ \cdot & -\alpha_3 v_{f_3} & \cdot & \cdot & A_{f_3} & \cdot & 0 & \cdot & \cdot & A_{f_3}^x \otimes 1 \end{array} \right] \cdot \quad (3.3)$$

$$\left[\begin{array}{ccccc} A_{11} & A_{12} & \cdot & \cdot & \cdot \\ A_{21} & \alpha_0 & u_{f_1} & u_{f_2} & u_{f_3} \\ \cdot & -\alpha_1 v_{f_1} & A_{f_1} & \cdot & \cdot \\ \cdot & -\alpha_2 v_{f_2} & \cdot & A_{f_2} & \cdot \\ \cdot & -\alpha_3 v_{f_3} & \cdot & \cdot & A_{f_3} \end{array} \right]$$

On the other hand, the system matrix of $\partial_{y_i|y'_i}\mathcal{A}_g$ is

$$\begin{bmatrix} A_{11} & A_{12} & A_{11}^{y_i} \otimes y'_i & A_{12}^{y_i} \otimes y'_i \\ A_{21} & a & A_{21}^{y_i} \otimes y'_i & \alpha_i y'_i \\ & & A_{11} & A_{12} \\ & & A_{21} & a \end{bmatrix},$$

the corresponding *partially* linearized system matrix of $(\partial_{y_i|y'_i}\mathcal{A}_g) \circ (\mathbf{f}, f'_i)$ is

$$\left[\begin{array}{ccccc|ccccc} A_{11} & A_{12} & \cdot & \cdot & \cdot & A_{11}^{y_i} \otimes f'_i & A_{12}^{y_i} \otimes f'_i & \cdot & \cdot & \cdot \\ A_{21} & \alpha_0 & u_{f_1} & u_{f_2} & u_{f_3} & A_{21}^{y_i} \otimes f'_i & \boxed{\alpha_i f'_i} & \cdot & \cdot & \cdot \\ \cdot & -\alpha_1 v_{f_1} & A_{f_1} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & -\alpha_2 v_{f_2} & \cdot & A_{f_2} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & -\alpha_3 v_{f_3} & \cdot & \cdot & A_{f_3} & \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline & & & & & A_{11} & A_{12} & \cdot & \cdot & \cdot \\ & & & & & A_{21} & \alpha_0 & u_{f_1} & u_{f_2} & u_{f_3} \\ & & & & & \cdot & -\alpha_1 v_{f_1} & A_{f_1} & \cdot & \cdot \\ & & & & & \cdot & -\alpha_2 v_{f_2} & \cdot & A_{f_2} & \cdot \\ & & & & & \cdot & -\alpha_3 v_{f_3} & \cdot & \cdot & A_{f_3} \end{array} \right].$$

To eliminate the boxed entry $\alpha_i f'_i$ we just need to recall the derivative $\partial_x \mathcal{A}_{f_i}$ of $\mathcal{A}_{f_i} = (u_{f_i}, A_{f_i}, v_{f_i})$, the invertible matrix Q swaps block columns 1 and 2:

$$\left[\begin{array}{c|cc} 0 & u_{f_i} & \cdot \\ \hline \cdot & A_{f_i} & A_{f_i}^x \otimes 1 \\ -\alpha_i v_{f_i} & \cdot & A_{f_i} \end{array} \right] Q = \left[\begin{array}{ccc} u_{f_i} & \boxed{0} & \cdot \\ A_{f_i} & \cdot & A_{f_i}^x \otimes 1 \\ \cdot & -\alpha_i v_{f_i} & A_{f_i} \end{array} \right].$$

Thus, after summing up (over $i \in \{1, 2, \dots, d\}$), we get the system matrix (3.3) and hence

$$\partial_x \mathcal{A}_h = \partial_x (\mathcal{A}_g \circ \mathcal{A}_f) = \partial_{\mathbf{y}|\mathbf{y}'} \mathcal{A}_g \circ \mathcal{A}_{\mathbf{f}, \mathbf{f}'},$$

that is, $\partial_x h = \partial_x (g \circ \mathbf{f}) = \partial_{\mathbf{y}|\partial_x \mathbf{f}} g \circ \mathbf{f}$. ■

Free (non-commutative) decomposition is important in *control theory* [12, Section 6.2.2]: “The authors do not know how to fully implement the decompose operation. Finding decompositions by hand can be facilitated with the use of certain type of collect command”.

Let $g = fabf + cf + de$ and $f = xy + z$, that is, f solves a Riccati equation. Given

$$h = g \circ f = (xy + z)ab(xy + z) + c(xy + z) + de$$

by the (minimal) admissible linear system $\mathcal{A}_h = (u_h, A_h, v_h)$,

$$\begin{bmatrix} 1 & -x & -z & \cdot & -d & -c & \cdot & \cdot \\ \cdot & 1 & -y & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & -a & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & -b & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & -e \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & -x & -z \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & -y \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix} s = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix},$$

one can read off $f = xy + z$ directly since A_h has the form

$$\begin{bmatrix} 1 & \boxed{-f} & \cdot & -d & -c & \cdot \\ \cdot & 1 & -a & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & -b & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & -e \\ \cdot & \cdot & \cdot & \cdot & 1 & \boxed{-f} \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix} s = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}.$$

So, starting with a *minimal* ALS $\tilde{\mathcal{A}}_h$ for h one “just” needs to find appropriate (invertible) transformation matrices P, Q such that $\mathcal{A}_h = P\tilde{\mathcal{A}}_hQ$ using (commutative) Gröbner bases techniques similar to [10, Theorem 4.1] or the refinement of pivot blocks [32, Section 3]. Although this is quite challenging using brute force methods, in practical examples it is rather easy using *free fractions*, that is, *minimal* and *refined* admissible linear systems, as a “work bench” where one can perform the necessary row and column operations manually.

“The challenge to computer algebra is to start with an expanded version of $h = g \circ f$, which is a mess that you would likely see in a computer algebra session, and to automatically motivate the selection of f .” [12, Section 6.2.1]

Working with *free fractions* is simple, in particular with nc polynomials. The main difficulty in working with (finite) formal power series is that the number of words can grow exponentially with respect to the *rank*, the minimal

dimension of a linear representation [31, Table 1]. In this case, for

$$h = g \circ f = xyabxy + xyabz + zabxy + abz + cxy + cz + de,$$

the main “structure” becomes (almost) visible already after *minimization* of the corresponding “polynomial” admissible linear system (in upper triangular form with ones in the diagonal).

4. APPLICATION: NEWTON ITERATION

To compute the third root $\sqrt[3]{z}$ of a (positive) real number z , say $z = 2$, we just have to find the roots of the polynomial $p = x^3 - z$, for example by using the Newton iteration $x_{k+1} := x_k - p(x_k)/p'(x_k)$, that —given a “good” starting value x_0 — yields a (quadratically) convergent sequence x_0, x_1, x_2, \dots such that $\lim_{k \rightarrow \infty} x_k^3 = z$. Using FRICAS [14], the first 6 iterations for $x_0 = 1$ are

k	$ x_k - x_{k-1} $	x_k
1	$3.333 \cdot 10^{-1}$	1.3333333333333333
2	$6.944 \cdot 10^{-2}$	1.2638888888888888
3	$3.955 \cdot 10^{-3}$	1.259933493449977
4	$1.244 \cdot 10^{-5}$	1.2599210500177698
5	$1.229 \cdot 10^{-10}$	1.2599210498948732
6	0	1.2599210498948732

Detailed discussions are available in every book on numerical analysis, for example [17]. As a starting point towards current research, one could take [33].

Now we would like to compute the third root $\sqrt[3]{Z}$ of a real (square) matrix Z (with positive eigenvalues). We (still) can use $X_{k+1} := \frac{2}{3}X_k + \frac{1}{3}X_k^{-2}Z$ if we choose a starting matrix X_0 such that $X_0Z = ZX_0$ because then all X_k commute with Z [19, Section 7.3]. For $X_0 = I$ and

$$Z = \begin{bmatrix} 47 & 84 & 54 \\ 42 & 116 & 99 \\ 9 & 33 & 32 \end{bmatrix} \quad \text{respectively} \quad \sqrt[3]{Z} = \begin{bmatrix} 3 & 2 & 0 \\ 1 & 4 & 3 \\ 0 & 1 & 2 \end{bmatrix},$$

some iterations are (with $\|\cdot\|_F$ denoting the Frobenius norm)

k	$\ X_k - X_{k-1}\ _F$	$\ X_k - \sqrt[3]{Z}\ _F$
4	9.874	23.679
5	6.511	13.818
6	4.109	7.309
7	2.254	3.200
8	$8.295 \cdot 10^{-1}$	$9.455 \cdot 10^{-1}$
9	$1.140 \cdot 10^{-1}$	$1.160 \cdot 10^{-1}$
10	$2.044 \cdot 10^{-3}$	$2.044 \cdot 10^{-3}$
11	$1.115 \cdot 10^{-6}$	$6.489 \cdot 10^{-7}$

Does this sequence of matrices X_0, X_1, \dots converge? Some more iterations reveal immediately that something goes wrong, visible in Table 1.

k	$\ X_k - X_{k-1}\ _F$	$\ X_k - \sqrt[3]{Z}\ _F$	$\ X_k Z - Z X_k\ _F$
1	65.836	5.385	$2.274 \cdot 10^{-13}$
2	22.279	60.756	$7.541 \cdot 10^{-13}$
3	14.845	38.500	$1.110 \cdot 10^{-12}$
4	9.874	23.679	$2.031 \cdot 10^{-12}$
5	6.511	13.818	$7.725 \cdot 10^{-12}$
6	4.109	7.309	$7.010 \cdot 10^{-11}$
7	2.254	3.200	$9.451 \cdot 10^{-10}$
8	$8.295 \cdot 10^{-1}$	$9.455 \cdot 10^{-1}$	$1.716 \cdot 10^{-8}$
9	$1.140 \cdot 10^{-1}$	$1.160 \cdot 10^{-1}$	$3.548 \cdot 10^{-7}$
10	$2.044 \cdot 10^{-3}$	$2.044 \cdot 10^{-3}$	$7.473 \cdot 10^{-6}$
11	$1.115 \cdot 10^{-6}$	$6.489 \cdot 10^{-7}$	$1.574 \cdot 10^{-4}$
12	$1.979 \cdot 10^{-5}$	$8.971 \cdot 10^{-7}$	$3.314 \cdot 10^{-3}$
13	$4.168 \cdot 10^{-4}$	$1.890 \cdot 10^{-5}$	$6.978 \cdot 10^{-2}$
14	$8.777 \cdot 10^{-3}$	$3.979 \cdot 10^{-4}$	1.469
15	$1.848 \cdot 10^{-1}$	$8.379 \cdot 10^{-3}$	$3.094 \cdot 10^{+1}$
16	3.892	$1.764 \cdot 10^{-1}$	$6.515 \cdot 10^{+2}$
17	$8.195 \cdot 10^{+1}$	3.715	$1.372 \cdot 10^{+4}$
18	$1.726 \cdot 10^{+3}$	$7.823 \cdot 10^{+1}$	$2.889 \cdot 10^{+5}$

Table 1: The first 18 Newton iterations to find $\sqrt[3]{Z}$ for $X_0 = I$. The values in column 2 (and 3) for the first 11 Newton steps seem to indicate convergence. However, the (Frobenius) norm of the *commutator* $X_k Z - Z X_k$ (column 4) increases steadily and that causes eventually a *diverging* sequence (X_k) .

The problem is that due to rounding errors (in finite precision arithmetics) commutativity of X_k (with Z) is lost, that is, $\|X_k Z - Z X_k\|_F$ increases with every iteration. This happens even for a starting value close to the solution $X_0 = \sqrt[3]{Z} + \varepsilon I$. For a detailed discussion of matrix roots and how to overcome such problems we refer to [19, Section 7], for further information about numerics to [11]. A classical introduction to matrix functions is [15, Chapter V]. For the (matrix) square root and discussions about the stability (of Newton's method) we recommend [18].

Remark. To ensure commutation of the iterates X_k with Z one could use an additional correction step solving the Sylvester equation $Z \Delta X_k - \Delta X_k Z = X_k Z - Z X_k$ for an update ΔX_k . Since this is expensive the benefit of using the “commutative” Newton iteration would be lost. For fast solutions of Sylvester equations we refer to [20].

Although what we are going to introduce now as “*non-commutative (nc) Newton iteration*” is even more expensive, it can be implemented as *black box* algorithm, that is, *without* manual computation of the non-commutative derivative and any individual implementation/programming. Furthermore, there is absolutely no restriction for the initial iterate.

To get the nc Newton method for solving $f(x) = 0$ we just need to “truncate” the Hausdorff *polarization operator* [27, Section 4]

$$f(x + b) = f(x) + \partial_{x|b} f(x) + \frac{1}{2} \partial_{x|b} (\partial_{x|b} f(x)) + \dots$$

as analogon to Taylor's formula. Thus, for $f(x) = x^3 - z$ we have to solve

$$0 = \underbrace{x^3 - z}_f + \underbrace{bx^2 + xbx + x^2b}_{\partial_{x|b} f} \in \mathbb{R}(\langle \mathcal{X} \rangle) =: \mathbb{F}$$

with respect to b . In terms of (square) matrices X, Z, B this is just the generalized Sylvester equation $BX^2 + XBX + X^2B = Z - X^3$ which is *linear* in B [19, Section 7.2]. Taking a (not necessarily with Z commuting) starting matrix X_0 , we can compute B_0 and get $X_1 := X_0 + B_0$ and iteratively $X_{k+1} := X_k + B_k$.

Minimal admissible linear systems for $f = x^3 - z$ and $\partial_{x|b} f = bx^2 + xbx + x^2b$ are given by

$$\begin{bmatrix} 1 & -x & \cdot & z \\ \cdot & 1 & -x & \cdot \\ \cdot & \cdot & 1 & -x \\ \cdot & \cdot & \cdot & 1 \end{bmatrix} s = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & -x & \cdot & -b & \cdot & \cdot \\ \cdot & 1 & -x & \cdot & -b & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & -b \\ \cdot & \cdot & \cdot & 1 & -x & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & -x \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix} s = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}$$

respectively. A *minimal* ALS $\mathcal{A} = (u, A, v)$ of dimension $n = 6$ for $g = f + \partial_{x|b}f$ is immediate:

$$\begin{bmatrix} 1 & -x & \cdot & -b & \cdot & z \\ \cdot & 1 & -x & \cdot & -b & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & -b-x \\ \cdot & \cdot & \cdot & 1 & -x & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & -x \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix} s = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}.$$

Since the system matrix is just a linear matrix pencil $A = A_1 \otimes 1 + A_b \otimes b + A_x \otimes x + A_z \otimes z$ we can plug in $m \times m$ matrices using the Kronecker product

$$A(B, X, Z) = A_1 \otimes I + A_b \otimes B + A_x \otimes X + A_z \otimes Z.$$

In this case, for $v = e_n = [0, \dots, 0, 1]^\top$, the evaluation of $g : M_m(\mathbb{R})^3 \rightarrow M_m(\mathbb{R})$ is just the upper right $m \times m$ block of $A(B, X, Z)^{-1}$. For how to efficiently evaluate (nc) polynomials by matrices using *Horner systems* we refer to [31].

Here, $A(B, X, Z)$ is invertible for arbitrary B, X and Z . Given Z and X , to find a B such that $g(B, X, Z) = 0$ we have to look for transformation matrices $P = P(T_{ij})$ and $Q = Q(U_{ij})$ of the form

$$P = \begin{bmatrix} I & \cdot & \cdot & T_{1,1} & T_{1,2} & 0 \\ \cdot & I & \cdot & T_{2,1} & T_{2,2} & 0 \\ \cdot & \cdot & I & T_{3,1} & T_{3,2} & 0 \\ \cdot & \cdot & \cdot & I & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & I & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & I \end{bmatrix}, \quad Q = \begin{bmatrix} I & \cdot & \cdot & 0 & 0 & 0 \\ \cdot & I & \cdot & U_{2,1} & U_{2,2} & U_{2,3} \\ \cdot & \cdot & I & U_{3,1} & U_{3,2} & U_{3,3} \\ \cdot & \cdot & \cdot & I & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & I & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & I \end{bmatrix}$$

such that the upper right block of size $3m \times 3m$ of $PA(B, X, Z)Q$ is zero, that is, we need to solve a *linear* system of equations (with $6m^2 + 6m^2 + m^2 = 13m^2$

unknowns) similar to the word problem [30, Theorem 2.4]. Table 2 shows the nc Newton iterations for the starting matrix

$$X_0 = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4.1)$$

using FRICAS [14] and the least squares solver DGELS from [22].

k	$\ B_k\ _F$	$\ X_k - \sqrt[3]{Z}\ _F$	$\ X_k Z - Z X_k\ _F$
0	46.877	5.745	113.842
1	16.081	42.298	5552.242
2	10.768	26.374	3659.788
3	7.320	15.912	2395.971
4	4.971	9.358	1534.892
5	2.934	6.122	912.700
6	2.651	4.389	414.201
7	1.380	1.846	86.771
8	$3.878 \cdot 10^{-1}$	$4.875 \cdot 10^{-1}$	5.638
9	$9.378 \cdot 10^{-2}$	$1.023 \cdot 10^{-1}$	$2.652 \cdot 10^{-1}$
10	$8.432 \cdot 10^{-3}$	$8.506 \cdot 10^{-3}$	$6.737 \cdot 10^{-3}$
11	$7.407 \cdot 10^{-5}$	$7.408 \cdot 10^{-5}$	$1.951 \cdot 10^{-5}$
12	$5.895 \cdot 10^{-9}$	$5.895 \cdot 10^{-9}$	$5.106 \cdot 10^{-10}$
13	$1.825 \cdot 10^{-14}$	$1.521 \cdot 10^{-14}$	$1.491 \cdot 10^{-12}$

Table 2: The first 13 (nc) Newton iterations to find $\sqrt[3]{Z}$ for X_0 from (4.1). In the beginning, the iterates X_k do not commute with Z (column 4).

In general, depending on the initial iterate X_0 , there is *no* guarantee that one ends up in some prescribed solution X since there can be several (matrix) roots. For the *principal* p -th root (and further discussion) we refer to [19, Theorem 7.2]. For a discussion of Taylor’s theorem (for matrix functions) one should have a look in [13].

The goal of this section was mainly for illustration (of the use of the free derivative). How could we attack analysis of convergence (for special classes of rational functions) in this context? “Classical” interval Newton is discussed in [28, Chapter 6.1]. What’s about “nc interval Newton”? The next natural

step would be *multi-dimensional* ne Newton, say to find a root $G(X, Y, Z) = 0$ for

$$G = \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} = \begin{bmatrix} x^3 + z - d \\ (x(1 - yx))^{-1} - e \\ xzx - yz + z^2 - f \end{bmatrix}$$

with (matrix-valued) parameters d, e, f . Although very technical, one can set up a “joint” linear system of equations to solve

$$\begin{pmatrix} g_1 + \partial_{x|a}g_1 & g_1 + \partial_{y|b}g_1 & g_1 + \partial_{z|c}g_1 \\ g_2 + \partial_{x|a}g_2 & g_2 + \partial_{y|b}g_2 & g_2 + \partial_{z|c}g_2 \\ g_3 + \partial_{x|a}g_3 & g_3 + \partial_{y|b}g_3 & g_3 + \partial_{z|c}g_3 \end{pmatrix} = 0$$

for matrices A, B and C using the previous approach to find tuples of “individual” transformation matrices P_{ij} and Q_{ij} to create respective upper right blocks of zeros. The problem however is, that this system is *overdetermined* in general, and starting arbitrary close to a root leads to a residual that causes divergence. How can one overcome that?

ACKNOWLEDGEMENTS

I am very grateful to Tobias Mai for making me aware of cyclic derivatives respectively the work of Rota, Sagan and Stein [27] in May 2017 in Graz and for a discussion in October 2018 in Saarbrücken. I also use this opportunity to thank Soumyashant Nayak for continuous “non-commutative” discussions, in particular about free associative algebras and (inner) derivations. And I thank the referees for the feedback and a hint on the literature.

REFERENCES

- [1] G.M. BERGMAN, W. DICKS, On universal derivations, *J. Algebra* **36**(2) (1975), 193–211.
- [2] J. BERSTEL, C. REUTENAUER, “Noncommutative Rational Series with Applications”, *Encyclopedia of Mathematics and its Applications*, 137, Cambridge University Press, Cambridge, 2011.
- [3] J.F. CAMINO, J.W. HELTON, R.E. SKELTON, Solving matrix inequalities whose unknowns are matrices, *SIAM J. Optim.* **17**(1) (2006), 1–36.
- [4] J.F. CAMINO, J.W. HELTON, R.E. SKELTON, J. YE, Matrix inequalities: a symbolic procedure to determine convexity automatically, *Integral Equations Operator Theory* **46**(4) (2003), 399–454.
- [5] P.M. COHN, Around Sylvester’s law of nullity, *Math. Sci.* **14**(2) (1989), 73–83.

- [6] P.M. COHN, “Skew Fields. Theory of General Division Rings”, *Encyclopedia of Mathematics and its Applications*, 57, Cambridge University Press, Cambridge, 1995.
- [7] P.M. COHN, “Further Algebra and Applications”, Springer-Verlag London, Ltd., London, 2003.
- [8] P.M. COHN, “Free Ideal Rings and Localization in General Rings”, *New Mathematical Monographs*, 3, Cambridge University Press, Cambridge, 2006.
- [9] P.M. COHN, C. REUTENAUER, A normal form in free fields, *Canad. J. Math.* **46** (3) (1994), 517–531.
- [10] P.M. COHN, C. REUTENAUER, On the construction of the free field, *Internat. J. Algebra Comput.* **9** (3-4) (1999), 307–323. Dedicated to the memory of Marcel-Paul Schützenberger.
- [11] J.W. DEMMEL, “Applied Numerical Linear Algebra”, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.
- [12] M.C. DE OLIVEIRA, J.W. HELTON, Computer algebra tailored to matrix inequalities in control, *Internat. J. Control* **79** (11) (2006), 1382–1400.
- [13] E. DEADMAN, S.D. RELTON, Taylor’s theorem for matrix functions with applications to condition number estimation, *Linear Algebra Appl.* **504** (2016), 354–371.
- [14] FRICAS TEAM, FriCAS — An advanced computer algebra system, 2019. Release 1.3.5, available at <http://fricas.sf.net>, documentation <http://fricas.github.io>.
- [15] F.R. GANTMACHER, “Matrizenrechnung. Teil I. Allgemeine Theorie”, Zweite, Berichtigte Auflage. Hochschulbücher für Mathematik, Band 36 VEB Deutscher Verlag der Wissenschaften, Berlin, 1965.
- [16] W. HACKBUSCH, “Hierarchische Matrizen: Algorithmen und Analysis”, Springer, Berlin, 2009.
- [17] P. HENRICI, “Elements of Numerical Analysis”, John Wiley & Sons, Inc., New York-London-Sydney, 1964.
- [18] N.J. HIGHAM, Newton’s method for the matrix square root, *Math. Comp.* **46** (174) (1986), 537–549.
- [19] N.J. HIGHAM, “Functions of Matrices. Theory and Computation”, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008.
- [20] P. KIRRINNIS, Fast algorithms for the Sylvester equation $AX - XB^T = C$, *Theoret. Comput. Sci.* **259** (1-2) (2001), 623–638.
- [21] D.S. KALIUZHNYI-VERBOVETSKYI, V. VINNIKOV, “Foundations of Free Noncommutative Function Theory”, *Mathematical Surveys and Monographs*, 199, American Mathematical Society, Providence, RI, 2014.
- [22] LAPACK 3.5.0, 2018, <http://www.netlib.org/lapack>.
- [23] G. POPESCU, Free holomorphic functions on the unit ball of $B(\mathcal{H})^n$, *J. Funct. Anal.* **241** (1) (2006), 268–333.
- [24] C. REUTENAUER, Cyclic derivation of noncommutative algebraic power series, *J. Algebra* **85** (1) (1983), 32–39.

- [25] C. REUTENAUER, Applications of a noncommutative Jacobian matrix, *J. Pure Appl. Algebra* **77** (2) (1992), 169–181.
- [26] C. REUTENAUER, Michel Fliess and non-commutative formal power series, *Internat. J. Control* **81** (3) (2008), 336–341.
- [27] G.-C. ROTA, B. SAGAN, P.R. STEIN, A cyclic derivative in noncommutative algebra, *J. Algebra* **64** (1) (1980), 54–75.
- [28] S.M. RUMP, Verification methods: rigorous results using floating-point arithmetic, *Acta Numer.* **19** (2010), 287–449.
- [29] K. SCHREMPF, Free fractions: An invitation to (applied) free fields, *ArXiv e-prints*, September 2018. Version 2, October 2020, <http://arxiv.org/pdf/1809.05425>.
- [30] K. SCHREMPF, Linearizing the word problem in (some) free fields, *Internat. J. Algebra Comput.* **28** (7) (2018), 1209–1230.
- [31] K. SCHREMPF, Horner Systems: How to efficiently evaluate non-commutative polynomials (by matrices), *arXiv e-prints*, October 2019.
- [32] K. SCHREMPF, A standard form in (some) free fields: How to construct minimal linear representations, *Open Math.* **18** (1) (2020), 1365–1386.
- [33] D. SCHLEICHER, R. STOLL, Newton’s method in practice: Finding all roots of polynomials of degree one million efficiently, *Theoret. Comput. Sci.* **681** (2017), 146–166.
- [34] D. VOICULESCU, A note on cyclic gradients, *Indiana Univ. Math. J.* **49** (3) (2000), 837–841.