# Pseudofactors of Multivariate Polynomials*


Yuzhen Huang
Wenda Wu
Beijing Municipal Computing Center
Beijing 100005, China
wdwu@mmrc.iss.ac.cn

Hans J. Stetter
Tech. Univ. Vienna
A-1040 Vienna, Austria
stetter@aurora.tuwien.ac.at

Lihong Zhi
Dept. of Computer Science
Ehime University
Matsuyama 790-8577, Japan
lzhi@hpc.cs.ehime-u.ac.jp



## ABSTRACT

We treat pseudo-factorization of a multivariate polynomial $p$ over $\mathbb{C}$ as an overdetermined bilinear system for the coefficients of the factors. If the specified data (coefficients of $p$) are sufficiently close to the manifold of *consistent* data we project onto it by solving a well-chosen subsystem (otherwise we quit). On the manifold, we reduce the backward error by a *linearized* minimization procedure. The resulting algorithm appears to extend the range of treatable cases – in terms of number of variables and total degree – considerably.


## 1. INTRODUCTION

When posed for polynomials over the complex or real numbers, several standard problems in polynomial algebra appear as *singular*: Solutions exist only for special polynomials, and an *arbitrarily small change* in their coefficients deletes the existence of a solution. One of these problems is the *factorization of a multivariate polynomial*: If $p \in \mathbb{C}[x_1, \dots, x_s]$, $s > 1$, happens to be factorizable, an infinitesimal generic perturbation of its coefficients will destroy this property. Due to this singular behavior, multivariate polynomials with some (non-integer) floating-point coefficients, as they may arise in the modelling of real-life situations, cannot be expected to be factorizable in the strict sense.

But factorizability exhibits an important structural property of a polynomial and its variety, with obvious consequences for the analysis of the underlying mathematical model. If the coefficientwise distance of the specified "empirical" polynomial from a factorizable polynomial is below the tolerance accuracy of the model, one may assume that the modelled situation has a factorizable structure. Such polynomials are often called *pseudo-factorizable* and the factors of a near-by factorizable polynomial *pseudofactors*. This pseudo-factorizability has not been a traditional subject of computer algebra (cf., e.g., [7],[4],[15]) where the interest is rather in (absolute) irreducibility and exact factors (e.g. [3] and many others); but it has gathered interest during the past decade because of the increasing use of computer algebra in Scientific Computing; cf. e.g. [10], [11], [6].


*Joint work under project V.A. 18 of the Chinese-Austrian Scientific-Technical Cooperation Agreement 1998/2000


A multivariate polynomial $p$ has a factor $u$ of a specified structure iff the coefficients of $u$ and of the associated factor $v$ in $p = u \cdot v$ satisfy a *system of bilinear equations*. Pseudosolutions of that *overdetermined* system correspond to pseudofactors of the polynomial $p$. Therefore, like other authors, we treat the search for pseudofactors as a special case of solving overdetermined polynomial systems approximately. The classical approach to that problem (cf., e.g., the seminal papers [2],[9],[6]) attempts to find solutions with the *shortest Euclidean residual*; this results in a *polynomial least squares problem*. But there may be many local minima, and the global minimum may still have a residual much too large. In contrast, we attempt to *separate the global and the local aspects* of the problem as indicated below.

In section 2, after the introduction of a "tolerance" metric in the coefficient space (cf. [14]), we introduce our algorithmic scheme for finding pseudozeros of overdetermined polynomial systems:

(i) Find candidates for pseudozeros from consistent subsystems;

(ii) Check and improve promising candidates

Phase (i) treats the global aspects of the problem in a conventional form. Phase (ii) is of a local nature so that *linearization* may be employed in the minimization; besides, there is no need to reach the actual minimum. If there are no pseudozeros, the computation will end prematurely.

In section 3, we turn to the special systems which arise from the multivariate factorization problem, with a *normalization* which appears advantageous. We exhibit the application of the above algorithmic scheme to that task with the aid of a simple example. In section 4, we consider the general case; we apply the same algorithmic pattern for the numerical determination of a pseudofactor with a specified support, for a general multivariate polynomial. Special cases requiring attention are pointed out and discussed. Finally, in section 5, we demonstrate the feasibility of our approach with a commented nontrivial example.

Naturally, our approach contains elements from some previous related work, e.g. [10], [11]; but we believe that our reduction to a local problem and our local techniques are original in this context.

# 2. PSEUDOZEROS OF OVERDETERMINED POLYNOMIAL SYSTEMS

## 2.1 Empirical Data and Pseudozeros

Mathematical models of real-life situations generally contain some *data of limited accuracy*: Each value $\tilde{\alpha}$ sufficiently close to the *specified value* $\bar{\alpha}$ is an *equally valid* specification of that data item.

**Definition 1**:[1] An *empirical data item* $\alpha = (\bar{\alpha}, \varepsilon)$, $\bar{\alpha} \in \mathbb{C}$, $\varepsilon > 0$, defines the *family of neighborhoods*

$$N_\delta(\bar{\alpha}, \varepsilon) := \{\tilde{\alpha} \in \mathbb{C} \; : \; \frac{|\tilde{\alpha} - \bar{\alpha}|}{\varepsilon} \le \delta\}\,, \; \delta \ge 0\,. \tag{1}$$

The *tolerance* $\varepsilon$ is assumed to express the indetermination in $\alpha$ such that values in $N_1(\bar{\alpha}, \varepsilon)$ may be considered as *indistinguishable.* But $\delta = 1$ is *not* a sharp limit but simply a mark on a continuous scale since $\varepsilon$ is generally known as an *order of magnitude* only. □

**Remark**: The last sentence is important: The (complex or real) *interval* $N_1(\bar{\alpha}, \varepsilon)$ is inadequate as a description of an empirical data item because it introduces an unrealistic *yes - no* situation; cf. the scale (5) below. Therefore, we have refrained from using the terminology and methodology of interval mathematics. □

Scalar empirical data may be freely combined into arrays or other data structures, in an obvious manner: E.g., an empirical data vector $a = (\bar{a}, e)$, $\bar{a} \in \mathbb{C}^n$, $e \in \mathbb{R}^n_+$, denotes the neighborhood family

$$N_\delta(\bar{a}, e) := \{\tilde{a} \in \mathbb{C}^n \; : \; \|(\ldots, \frac{|\tilde{\alpha}_\nu - \bar{\alpha}_\nu|}{\varepsilon_\nu}, \ldots)\|^* \le \delta\}\,, \tag{2}$$

where $\|\ldots\|^*$ is some fixed absolute norm in $\mathbb{R}^n$. In particular, we will deal with empirical polynomials:

**Definition 2**: An *empirical polynomial* $p = (\bar{p}, e)$, with $\bar{p} = \sum_{j \in J} \bar{\alpha}_j x^j \in \mathbb{C}[x]$, $x = (x_1 \ldots, x_s)$, $s \ge 1$, $\bar{\alpha}_j \in \mathbb{C}$, $J \subset \mathbb{N}_0^s$, and $e = (\varepsilon_j, j \in J), \varepsilon_j \ge 0$, defines a *family of polynomial neighborhoods* $(\delta > 0)$

$$\begin{aligned} N_\delta(\bar{p}, e) := \; & \{\tilde{p} \in \mathbb{C}[x] \; : \\ \tilde{p} = & \sum_{j \in J} \tilde{\alpha}_j x^j, \; \|(\ldots, \tfrac{|\tilde{\alpha}_j - \bar{\alpha}_j|}{\varepsilon_j}, \ldots)\|^* \le \delta\}\,. \end{aligned} \tag{3}$$

The polynomials in $N_1(\bar{p}, e)$ will be considered as *indistinguishable.* □

Note that all polynomials in $N_\delta(\bar{p}, e)$ have the *same support* $J$. We may admit $\varepsilon_j = 0$ for some $j \in J$ to accomodate *intrinsic* coefficients which are known exactly (e.g. integer ones). For such $j$, $\tilde{\alpha}_j = \bar{\alpha}_j$ for all $\tilde{p} \in N_\delta(\bar{p}, e)$. Via (3) and (2), the neighborhoods $N_\delta(\bar{p}, e)$ define a *distance from* $\bar{a}$ in the coefficient space $\mathcal{A}$ which we denote by $\|a - \bar{a}\|_e^*$.

A system $P = (\bar{P}, E)$ of empirical polynomials may be considered as a vector of empirical polynomials in an obvious manner. Note that this introduces a further choice of a vector norm through (2). Algebraic problems are called empirical if some of their data are empirical. In particular, the empirical system $(\bar{P}, E)$ defines an *empirical system of polynomial equations* $\bar{P}(x) = 0$, with tolerance $E$.

**Definition 3**: $\tilde{z} \in \mathbb{C}^s$ is a $\delta$*-pseudozero* of the empirical system $(\bar{P}, E) = \{(\bar{p}_\nu, e_\nu), \nu = 1(1)n\}$, $\bar{p}_\nu \in \mathbb{C}[x_1, \ldots, x_s]$, if there exists a system $\tilde{P} \in N_\delta(\bar{P}, E)$ such that $\tilde{z}$ is a *genuine* zero of $\tilde{P}$. A 1-pseudozero is simply called a *valid* pseudozero of $\bar{P}$. The $\delta$*-pseudozero set* is

$$Z_\delta(\bar{P}, E) := \{\tilde{z} \; : \; \tilde{z} \text{ is a } \delta\text{-pseudozero of } (\bar{P}, E)\}. \quad \Box \tag{4}$$

The $\delta$-pseudozero sets $Z_\delta$ with $\delta > 1$ may be viewed as sets of approximate zeros with less and less validity as indicated by some scale like

| $\delta \le 1$ | $\ldots$ 3 | $\ldots$ 10 | $\ldots$ 30 | $\ldots$ 100 | |
|---|---|---|---|---|---|
| valid | probably valid | possibly valid | probably invalid | invalid | (5) |

When the (exact) zero set $Z_0(\bar{P})$ of the specified system $\bar{P}$ consists of $m$ well isolated zeros, $Z_\delta(\bar{P}, E)$ for $\delta > 0$ generally consist of $m$ separate connected and bounded domains, the $\delta$*-pseudozero domains.* A pseudozero domain with several exact zeros of $\bar{P}$ (counting multiplicities) defines a *zero cluster* relative to the tolerance $E$.

## 2.2 Backward Error of Pseudozeros

All these definitions would be of academic value only, if one could not decide the membership of some $\tilde{z}$ in $Z_\delta$ computationally.

**Definition 4**: In the space $\mathcal{A}$ of the empirical coefficients of an empirical polynmomial $(\bar{p}, e)$ or system $(\bar{P}, E)$, for a specified "approximate zero" $\tilde{z}$, the manifold

$$\begin{aligned} \mathcal{M}(\tilde{z}) := \{ & \tilde{\alpha}_j, j \in J \; : \\ & \tilde{z} \text{ is an } \textit{exact} \text{ zero of } \tilde{p} \text{ or } \tilde{P} \text{ with coefficients } \tilde{\alpha}_j\} \end{aligned} \tag{6}$$

is the *equivalent-data manifold* for $\tilde{z}$. □

**Definition 5**: For a specified approximate zero $\tilde{z}$ of an empirical polynomial $(\bar{p}, e)$, its *backward error* is given by

$$d(\tilde{z}) = d(\tilde{z}; (\bar{p}, e)) := \min_{a \in \mathcal{M}(\tilde{z})} \|a - \bar{a}\|_e^*\,. \tag{7}$$

For a system $(\bar{P}, E)$, the definition is analogous. □

**Proposition 1**: $\tilde{z}$ is a $\delta$-pseudozero of $(\bar{p}, e)$ or $(\bar{P}, E)$ resp. iff

$$d(\tilde{z}; (\bar{p}, e)) \;\text{ or }\; d(\tilde{z}; (\bar{P}, E)) \;\le\; \delta\,.$$

**Remark**: In most of the relevant literature (cf. e.g. [1]), no weighting by tolerances is used in the definition of the backward error. This implies that the corresponding Proposition 1 must require that the backward error is *below the tolerance level.* We have found that our scaling provides a welcome flexibility for data with different tolerance levels.

[1] Naturally, $\mathbb{C}$ may be replaced by $\mathbb{R}$ in the following

*Example*: Consider the univariate empirical polynomial $p = (\bar{p}, e)$, $\bar{p} = \sum_{j=0}^{n} \bar{\alpha}_j x^j$. For $\tilde{z} \in \mathbb{C}$,

$$\mathcal{M}(\tilde{z}) = \{a \in \mathbb{C}^{n+1} : \sum_{j=0}^{n} (\alpha_j - \bar{\alpha}_j)\tilde{z}^j + \bar{p}(\tilde{z}) = 0\}$$

which is a *linear manifold* of codimension 1 in $\mathcal{A} = \mathbb{C}^{n+1}$. Its minimal $e$-distance from $\bar{a}$ is given by (cf. e.g. [1],[14])

$$d(\tilde{z}; (\bar{p}, e)) = \begin{cases} \frac{|\bar{p}(\tilde{z})|}{\sum \varepsilon_j |\tilde{z}|^j} & \text{for max-norm in (3)} \\ \frac{|\bar{p}(\tilde{z})|}{\sqrt{\sum (\varepsilon_j |\tilde{z}|^j)^2}} & \text{for 2-norm in (3)} \end{cases} \qquad (8)$$

By Definition 2, these expressions remain *unchanged* for a *multivariate* empirical polynomial $(\bar{p}, e)$, with $\bar{p} \in \mathbb{C}[x_1..x_s]$, and an approximate zero $\tilde{z} \in \mathbb{C}^s$; only the summation in $\mathcal{M}(\tilde{z})$ and $d(\tilde{z})$ now runs over the empirical coefficients $\alpha_j$ with $j \in J \in \mathbb{N}_0^s$ (cf. [14]). □

If $\mathcal{M}(\tilde{z})$ is not a linear manifold of codimension 1, there are generally no explicit expressions for $d(\tilde{z})$; but there are well-known algorithms for the evaluation of (7) with the commonly used norms.

For a system, one would like to request that

$$Z_\delta(\bar{P}, E) = \cap_\nu Z_\delta(\bar{p}_\nu, e_\nu), \qquad (9)$$

i.e. that a $\delta$-pseudozero of each $(\bar{p}_\nu, e_\nu)$ is a $\delta$-pseudozero of $(\bar{P}, E)$ . This requires the use the max-norm for the vector of polynomials $p_\nu$ in the neighborhood definition for $(\bar{P}, E)$.

If $\bar{P}$ is an *overdetermined system* it will generally not possess genuine zeros and $Z_0(\bar{P}) = \cap_{\delta>0} Z_\delta(\bar{P}, E) = \emptyset$. However, in a set topology, the individual $Z_\delta(\bar{p}_\nu, e_\nu)$ increase monotonously with $\delta$; they cover the whole result space for $\delta \to \infty$. Thus, by (9), $Z_\delta(\bar{P}, E) \neq \emptyset$ for sufficiently large $\delta$ which implies the existence of $\underline{\delta} := \inf\{\delta \ : \ Z_\delta(\bar{P}, E) \neq \emptyset\}$. $(\bar{P}, E)$ has valid pseudozeros iff $\underline{\delta} \leq 1$.

## 2.3 Pseudozero Computation for Overdetermined Systems

Given the overdetermined empirical system $\bar{Q} = \{\bar{q}_\nu, \nu = 1(1)n\}$ with tolerance $E$ in $s < n$ variables, we want to find some $\tilde{z} \in \mathbb{C}^s$ with

$$d(\tilde{z}; (\bar{Q}, E)) \ = \ \max_\nu d(\tilde{z}; (\bar{q}_\nu, e_\nu)) \ \leq \ 1\,, \qquad (10)$$

with the max-norm for the $(\bar{q}_\nu, e_\nu)$ in (3). For a realistic situation, we assume that the $\varepsilon_{\nu j}$ in the $e_\nu$ are small and that we are interested in pseudozeros with a moderate $|\tilde{z}|$, i.e. $|\tilde{z}| = O(1)$.

### (i) Determination of Candidates for Pseudozeros

By (8), a $\delta$-pseudozero $\tilde{z}$ of $(\bar{Q}, E)$ must satisfy

$$|\bar{q}_\nu(\tilde{z})| \ \leq \ \delta \sum_{j \in J_\nu} \varepsilon_{\nu j} |\tilde{z}|^j\,, \quad \nu = 1(1)n\,, \qquad (11)$$

If we select a suitable *consistent* subset $\bar{Q}_0 = \{\bar{q}_{\nu_\sigma}, \sigma = 1(1)s\}$ of $\bar{Q}$, its *zero set* $Z(\bar{Q}_0)$ should contain potential candidates for pseudozeros of $(\bar{Q}, E)$. More generally, we can use any consistent set of necessary conditions for a zero of $\bar{Q}$ for the determination of a candidate. Such a set may contain *fewer than s* variables which may diminish the effort for finding approximate solutions; with these components, the remaining components of the candidates may be obtained from further necessary conditions.

This strategy permits the use of the algebraic machinery for the reduction of the original *global* problem to a *local* one which is the principal difficulty of our task. The actual selection of $\bar{Q}_0$ will be guided by special properties of $\bar{Q}$ as shown in section 3 for the case of mulitvariate factorization.

### (ii) Checking and Improvement of Candidate Pseudozeros

For a candidate pseudozero $\tilde{z}_0$, from whatever source, we compute the backward errors $d_\nu(\tilde{z}_0) := d(\tilde{z}_0; (\bar{q}_\nu, e_\nu)), \nu = 1(1)n$; cf. (10). If $\tilde{z}_o$ has been found as an approximate zero of the consistent polynomial system $\bar{Q}_0 \subset \bar{Q}$, we have $d_{\nu_\sigma}(\tilde{z}_0) \ll 1$ for $\sigma = 1(1)s$ and it suffices to compute the remaining $d_\nu(\tilde{z}_0)$. There are three possibilities (cf. (10)):

(1) $d(\tilde{z}_0; (\bar{Q}, E)) \ \leq \ 1$ : $\tilde{z}_0$ is a *valid zero* of $(\bar{Q}, E)$ and we are either finished (if one pseudozero is sufficient) or we may proceed to further candidates.

(ii) $d(\tilde{z}_0; (\bar{Q}, E)) \ > \ 1$ but within reasonable bounds : $\tilde{z}_0$ is not a valid approximate zero and we will *correct* it in order to decrease $d(\tilde{z}_0)$.

(iii) $d(\tilde{z}_0; (\bar{Q}, E)) \ \gg \ 1$ : $\tilde{z}_0$ is not a suitable candidate.

Note that we need only a very *low relative accuracy* in the evaluation of (8) for the purpose of this stage.

Now consider the system of $n > s$ empirical polynomials $(\bar{q}_\nu, e_\nu)$ in $s$ variables and an approximate zero $\tilde{z}_0 \in \mathbb{C}^s$ which is not valid, i.e. $\max_\nu d_\nu(\tilde{z}_0) > 1$. From (8) we know that $d_\nu(z)$ is a weighted modulus of the residual $\bar{q}_\nu(z)$. If we proceed from $\tilde{z}_0$ to $\tilde{z}_0 - \Delta z$, with $|\Delta z| \ll |\tilde{z}_0|$, the relative and absolute change in the denominator in (8) will be small while the relative change of the residual will be large if $|p_\nu(\tilde{z}_0)|$ is small. Therefore, in a local linearization of $d_\nu$ about $\tilde{z}_0$, it is natural to consider the weights as constants and to linearize only the residual. Hence, we replace $d_\nu(\tilde{z}_0 - \Delta z)$ by

$$\tilde{d}_{\nu 0}(\Delta z) \ := \ \frac{1}{w_\nu(\tilde{z}_0)} \, |\bar{q}_\nu(\tilde{z}_0) - (\text{grad } \bar{q}_\nu(\tilde{z}_0))\,\Delta z|\,, \qquad (12)$$

with $w_\nu(z) := \sum_{j \in J_\nu} \varepsilon_{\nu j} |z|^j$, and we determine the *correction* $\Delta z$ from

$$\min_{\Delta z} \ \max_\nu \ \tilde{d}_{\nu 0}(\Delta z)\,. \qquad (13)$$

Thus, our approach has replaced the *global nonlinear nonconvex* minimization problem

$$\min_z \ \max_\nu \ d_\nu(z) \qquad (14)$$

by the *local linear* or *convex* minimization problem (13).

If the coefficients and the solution are restricted to be *real*, (13) is a standard *linear minimization* problem; it may be written as a linear program and solved with standard software. (If the solution is not unique, it is good numerical practice to choose the correction with the smallest norm.)

In the *complex* case, the $\tilde{d}_{\nu 0}(\Delta z)$ of (12) are no longer piecewise linear but they remain *convex* functions so that (13) has a solution and can be solved by standard minimization software. The correction step (13) is the well-known max-norm counterpart to the Gauss-Newton correction step for the Euclidean norm

$$\min_{\Delta z} \sum_{\nu} [\tilde{d}_{\nu 0}(\Delta z)]^2 .$$

It is not our intention to use the correction step (13) in an iterative fashion: If $\tilde{z}_0$ is a suitable candidate (cf. section 2.3.2), one step (13) should generally suffice to make $d(\tilde{z}_0 - \Delta z)$ sufficiently small. If $d(\tilde{z}_0 - \Delta z) \approx d(\tilde{z}_0)$ but both are too big, we must conclude that we are close to a local minimum which does not represent a valid pseudozero of $(\bar{Q}, E)$. Note that we do not search for a minimum of $d(z)$ but only for a point $z$ where $d(z)$ is sufficiently small.

A more abstract description of the situation is the following: In the data space $\mathcal{A}$, the coefficient sets $a$ of *consistent* systems $P$ define an algebraic variety $\mathcal{C}$. Our specified system $\bar{Q}$ has coefficients $\bar{a} \notin \mathcal{C}$. Via (7), the global optimization problem (14) defines a projection from (almost all of) $\mathcal{A}$ onto $\mathcal{C}$ which is, generally, very difficult to realize algorithmically.

Our candidate search (i) implicitly defines a mapping from the *neighborhood* of $\mathcal{C}$ in $\mathcal{A}$ onto $\mathcal{C}$. In the correction phase (ii), we move *along* $\mathcal{C}$ towards a point with a smaller value of $d$. If $\bar{a}$ has been sufficiently close to $\mathcal{C}$, this results in a satisfactory solution. Clearly, this is a more pragmatic approach; but it is well suited for the applications which we have described in the Introduction. In this formulation, our approach is seen to be related to that of [5], both in its objective and in some principal ideas; but our techniques are different.

# 3. PSEUDOFACTORIZATION OF MULTIVARIATE POLYNOMIALS

Consider a polynomial $p = \sum_{j \in J} \alpha_j x^j \in \mathbb{C}[x_1, \ldots, x_s]$ of total degree $k$, with support $J \in N_0^s$. A factorization of $p$ into two factors $u$ of total degree $l$ and $v$ of total degree $k - l \geq l$ can exist only if the *overdetermined system of equations* for the coefficients $\beta_j$ of $u$ and $\gamma_j$ of $v$ which results from a comparison of coefficients in

$$p(x; a) \;:=\; \sum_{j \in J} \alpha_j \, x^j \;=\; \Big(\sum_{|j| \leq l} \beta_j \, x^j\Big) \Big(\sum_{|j| \leq k-l} \gamma_j \, x^j\Big) \qquad (15)$$

has at least one solution. Note that, generally, (15) will contain *all* equations for $|j| \leq k$ including those with $j \notin J$ where $\alpha_j = 0$; only for special sparse supports, corresponding sparse supports for $u, v$ may be chosen directly (cf. also section 4.2.1).

Exact factorizability is a very restrictive requirement for a multivariate polynomial. E.g., for $s = 3$, $k = 5$, $l = 2$, there are 56 equations for the 30 coefficients of $u$ and $v$. However, if the real-world phenomenon which has been modelled by $p$ is decomposible, we should be able to find pseudofactors of $p$, i.e. factors of some factorizable polynomial $\tilde{p} \in N_\delta(\bar{p}, e)$, with $\delta = \mathrm{O}(1)$. This task is the main objective of this paper.

Obviously, we may *normalize* one coefficient in each of the 3 polynomials in (15). If $p$ possesses a nonvanishing constant term $\alpha_0$, it appears convenient to require

$$\alpha_0 \;=\; \beta_0 \;=\; \gamma_0 \;=\; 1 \, . \qquad (16)$$

This is also the only simple normalization impartial with respect to the variables.

If $|\alpha_0| = 0$ or very small, (16) cannot be used; we will not consider the adaptation of the following considerations to such cases in this paper but assume (16) throughout. We are also aware that (16) deviates from the more customary way of normalizing some "leading" coefficients (cf. e.g. [6]); but the symmetry in the variables and the fixed constant term are important formal advantages. Clearly, our approach is in no essential way dependent on this normalization.

As in section 2, we do not require that *all* coefficients $(\bar{\alpha}_j, \varepsilon_j)$ of the given empirical polynomial $(\bar{p}, e)$ have a *positive* tolerance $\varepsilon_j$. In particular, $\bar{\alpha}_0 = 1$ will always be considered as exact; vanishing coefficients of $\bar{p}$ may or may not be designated as intrinsic.

## 3.1 The Case $s = 2$, $l = 1$

We use this simplest case for an *exposition* of our approach in which we also neglect special cases which may arise. From a comparison of coefficients in

$$\begin{aligned} p(x, y; \bar{a}) \;&=\; 1 + \sum_{1 \leq j_1 + j_2 \leq k} \bar{\alpha}_{j_1 j_2} \, x^{j_1} \, y^{j_2} \;= \\ &(1 + \beta_{10} \, x + \beta_{01} \, y)\,\Big(1 + \sum_{1 \leq j_1 + j_2 \leq k-1} \gamma_{j_1 j_2} \, x^{j_1} \, y^{j_2}\Big), \end{aligned} \qquad (17)$$

we obtain the overdetermined system

$$Q(b, c; \bar{a}) \;=\; \{q_{j_1 j_2}(b, c; \bar{a}),\; 1 \leq j_1 + j_2 \leq k\} \;=\; 0 \qquad (18)$$

for $b, c$ with $\binom{k+2}{2} - 1 = (k^2 + 3k)/2$ *bilinear* equations for the $2 + \binom{k+1}{2} - 1 = (k^2 + k + 2)/2$ variables $\beta_{10}, \beta_{01}$ and $\gamma_{j_1 j_2}$. Through the empirical coefficients of $(\bar{p}, e)$, (18) is an empirical system of which we want to find potential pseudozeros.

### (i) Determination of Candidate Coefficients

When an exact multivariate factorization problem is posed over $\mathbb{Q}$ (or over a finite field), it is common practice (cf. e.g. [4]) to consider the *univariate factorization problem* for (say) $p(x, 0)$. Under certain conditions, if this polynomial is irreducible, then $p(x, y)$ is also irreducible; also, there are various techniques to extend factors of $p(x, 0)$ into potential factors of $p(x, y)$ as has been widely discussed in the literature. Over $\mathbb{C}$, on the other hand, *all univariate polynomials factor* and $p(x, 0)$ has $k$ linear factors, $\binom{k}{2}$ quadratic factors etc. Thus, it is our first task to test which of these factors – if any – may lead to a multivariate pseudofactor. This strategy has also been suggested and used in [12].

Computationally, we form the univariate (approximate) factorization

$$p(x, 0; \bar{a}) \;\approx\; \prod_{\kappa=1}^{k} (1 + \beta_{10,\kappa} x) \, , \quad \beta_{10,\kappa} \in \mathbb{C} \, , \qquad (19)$$

with $\beta_{10,\kappa} = -1/\xi_\kappa$, where the $\xi_\kappa \neq 0$ (cf. (16)) are the $k$ approximate zeros of $p(x, 0; \bar{a})$. For each $\beta_{10,\kappa}$, we compute a unique extension $1 + \beta_{10,\kappa} x + \beta_{01,\kappa} y$ of the "germ" $1 +$

$\beta_{10,\kappa} x$. By the symmetry in the variables, $-1/\beta_{01,\kappa}$ should be a pseudozero of $p(0,y;\bar{a})$, which is tested by (3). For the extension, we consider the linear terms in the development of $p(x) - u(x)\,v(x)$ in powers of $y$ :

$$\begin{aligned} & p_0(x) + p_1(x)y + \ldots \;= \\ & (1 + \beta_{10}x + \beta_{01}y)\;(v_0(x) + v_1(x)y + \ldots)\,. \end{aligned} \tag{20}$$

With $\beta_{10,\kappa} = -1/\xi_\kappa$, $x = \xi_\kappa$, this yields $-1/\beta_{01,\kappa} = -\,v_0(\xi_\kappa)/p_1(\xi_\kappa)$ for the test in $p(0,y;\bar{a})$.

Iff this test is passed, we extend $b_0 = (\beta_{10}, \beta_{01})$ to a full set $b_0, c_0$ of candidate coefficients for $u$ and $v$ by further consideration of the terms in (20). This yields triangular systems, with a diagonal of 1's, for the coefficients of $v_1, v_2, \ldots$ and defines the coefficients $c_0$ of $v$.

#### (ii) Checking and Correction of Candidate Coefficients

In this fashion, we have used all equations of the overdetermined system (18) except the $q_j$, $j = (k-\kappa, \kappa), \kappa = 2(1)k$. As is obvious from (17), each of these contains just one data coefficient $\alpha_j$, with tolerance $\varepsilon_j$. Thus the backward error of a candidate pair $b_0, c_0$ has the nontrivial components

$$d_{k-\kappa,\kappa}(b_0, c_0) \;=\; \frac{|q_{k-\kappa,\kappa}(b_0, c_0; \bar{a})|}{\varepsilon_{k-\kappa,\kappa}}\,, \quad \kappa = 2(1)k\,, \tag{21}$$

for $\varepsilon_{k-\kappa,\kappa} \neq 0$; for $\varepsilon_{k-\kappa,\kappa} = 0$, it is not defined and we must turn to the correction stage.

If $\max_\kappa d_{k-\kappa,\kappa}(b_0,c_0) < 1$, we have found a *valid factorization* (pseudo-factorization) of the empirical polynomial $(\bar{p}, e)$. Otherwise, we turn to the correction stage – except if $\max_\kappa d_{k-\kappa,\kappa}(b_0, c_0)$ is very large so that we choose to discard the candidate pair.

For the *correction* of a candidate pair $b_0, c_0$, we follow section 2.3 : We linearize the complete overdetermined system (18) about $b_0, c_0$ and form the *linearized backward error* $\tilde{d}(\Delta b, \Delta c, b_0, c_0; \bar{a})$; cf. (12). Due to the special structure of (18), the *linear minimization problem* (13) becomes

$$\begin{aligned} & \min_{\Delta b, \Delta c}\; \tilde{d}(\Delta b, \Delta c, b_0, c_0; \bar{a}) \;:= \\ & \min_{\Delta b, \Delta c} \max_{1 \le |j| \le k} \frac{|q_j(b_0,c_0;\bar{a}) - q_j(\Delta b, c_0; \bar{a}) - q_j(b_0, \Delta c; \bar{a})|}{\varepsilon_j}\,. \end{aligned} \tag{22}$$

The occurrence of an $\varepsilon_j = 0$ leads to an equality condition $|q_j(b_0,c_0;\bar{a}) - q_j(\Delta b, c_0; \bar{a}) - q_j(b_0, \Delta c; \bar{a})| = 0$ ; instead, we may rather introduce a suitably chosen *small* $\varepsilon_j > 0$, at least in the first correction step.

Due to the fact that we have *linearized* our problem for the correction step, we have to reevaluate $Q(b_0 - \Delta b, c_0 - \Delta c; \bar{a})$; the decisions to be taken depend on the backward error obtained.

### 3.2 A Simple Example

Consider the empirical bivariate polynomial $(\bar{p}, e)$ with

$$\begin{aligned} \bar{p}(x,y) = {} & 1 - .82\,x + .91\,y + .15\,x^2 + 11.22\,xy - 8.71\,y^2 \\ & + 4.69\,x^3 - .65\,x^2 y - 12.08\,xy^2 + 7.14\,y^3 \end{aligned}$$

and all tolerances $\varepsilon_j = .005$. $\bar{p}(x,0) = 1 - .82\,x + .15\,x^2 + 4.69\,x^3$ has the zeros $\xi_1 \approx -.70655$ , $\xi_{2,3} \approx .33728 \pm .43361$ i. $\xi_1$ yields $\beta_{10} = -1/\xi \approx 1.41533$ and $-1/\beta_{01} \approx .57663$, with a max-norm backward error $.44 < 1$ in $\bar{p}(0,y)$. For $\xi_{2,3}$, on the other hand, the corresponding backward error becomes $\approx 800$ so that we abandon this choice and continue with $(\beta_{10}, \beta_{01})$.

From (20), we obtain $(\gamma_{10}, \gamma_{20}, \gamma_{01}, \gamma_{11}, \gamma_{02}) \approx (-2.23533,$ $2.64421, 3.31373, 3.60105, -4.12440)$. When we enter our candidate coefficients into the overdetermined system (18), we obtain a backward error of $\approx 2.5$. We may either accept that (cf. (5)) or form the minimization problem (22): There are the 7 components of $\Delta b, \Delta c$ to minimize the max of 9 linear expressions. It turns out that we can reduce the backward error of (18) below 1 by non-zero choices of $\Delta\beta_{01}$ and $\Delta\gamma_{02}$, leaving all other components of $b_0$ $c_0$ unchanged. The resulting pseudofactors

$$\begin{aligned} & \tilde{u}(x,y) = 1 + 1.4153\,x - 1.7330\,y\,, \quad \tilde{v}(x,y) = \\ & 1 - 2.2353\,x + 2.6442\,y + 3.3137\,x^2 + 3.6011\,xy - 4.1228\,y^2. \end{aligned}$$

satisfy

$$\begin{aligned} & p(x,y) - \tilde{u}(x,y)\,\tilde{v}(x,y) \approx -.0012\,y - .0001\,x^2 + .0028\,xy \\ & \quad - .0048y^2 + .0001\,x^3 - .0040\,x^2y - .0043\,xy^2 - .0048\,y^3\,, \end{aligned}$$

which establishes the pseudo-factorizability of $(\bar{p}, e)$. Note that these pseudofactors are just one pair from a dense set of pseudofactors, and that they do not represent the exact minimum of the backward error. But – under our assumptions – *all* pseudofactors with a backward error $\le 1$ are *equally valid* because the polynomials in $N_1(\bar{p}, e)$ *cannot be distinguished.* Thus, there are no "better" or "worse" pseudofactors.

## 4. THE GENERAL CASE

### 4.1 The Algorithmic Scheme

We now consider the (pseudo)factorizability of an $s$-variate polynomial $p \in \mathbb{C}[x]$ of total degree $k$, with constant terms 1 in $p$ and in the two factors $u$ of degree $l \le k/2$ and $v$ of degree $k - l$; cf. (15). At first, we assume a *generic* case; special situations will be considered in section 4.2. We assume that $p = (\bar{p}, e)$ is empirical, with tolerances $e$; cf. Definition 2. (Pseudo)factorizability is equivalent to (pseudo)solvability of the bilinear overdetermined system (cf. (18))

$$Q(b,c;\bar{a}) \;=\; \{q_j(b,c;\bar{a}),\; j \in \mathbb{N}_0^s, 1 \le |j| \le k\} \;=\; 0\,, \tag{23}$$

obtained by a comparison of coefficients in (15). Algorithmically, we follow the approach described in section 3.1 for $s = 2$ and $l = 1$.

#### 4.1.1 Quick Rejection of Unsuitable Germs

Without loss of generality, we use $x_1$ as the germ variable and $x_2$ as the extension variable; we disregard the remaining variables which are asssigned the value 0 in the present context. Consider an $l$-degree germ $\hat{u} = 1 + \beta_{10}x_1 + \cdots + \beta_{l0}x_1^l = \prod_{\lambda=1}^{l}(1 - \frac{x_1}{\xi_{\kappa_\lambda}})$, where the $\xi_{\kappa_\lambda}$ are $l$ of the zeros of $p(x_1, 0; \bar{a})$. Since there are $\binom{k}{l}$ such germs, with a high probability of unsuitability, the test should require a minimal effort per germ. We use the

**Observation**: $\prod_{\kappa=1}^{k}(1+a_\kappa y) = (1+b_1 y+\cdots+b_l y^l)\times(1+c_1 y+\cdots+c_{k-l}y^{k-l})$ *only if* $b_1 = \sum_{\lambda=1}^{l} a_{\kappa_\lambda}$ for some selection of $l$ $a_\kappa$'s. □

Therefore, we compute all such target values for $p(0,x_2)$ and check the extension coefficient $\beta_{01}$ of $x_2$ for the germ $\hat{u}$ against it, on the basis of the specified tolerances.

We explain the computation of $\beta_{01}$ for $l=2$, from which it is clear that it is also feasible for larger $l$. Let

$$p(x_1,x_2;\bar{a}) = p_{10}(x_1)+p_{11}(x_1)\,x_2+p_{12}(x_1)\,x_2^2+\cdots = (\hat{u}(x_1)+(\beta_{01}+\beta_{11}x_1)\,x_2+\beta_{02}x_2^2)\,(v_{10}(x_1)+v_{11}(x_1)\,x_2+v_{12}(x_1)\,x_2^2+\ldots)\,; \tag{24}$$

Since the germ $\hat{u}$ is a factor of $p_{10}$, we also know $v_{10} = p_{10}/\hat{u}$; we assume that $v_{10}$ and $\hat{u}$ have no common zeros, i.e. multiple zeros of $p_{10}$ have not been split between $\hat{u}$ and $v_{10}$. Comparison of the $x_2$-terms in (24) yields

$$(\beta_{01}+\beta_{11}x_1)\,v_{10}(x_1)+\hat{u}(x_1)\,v_{11}(x_1) = p_{11}(x_1)\,. \tag{25}$$

Substitution of the two zeros of $\hat{u}$ into (25) gives two linear equations for $\beta_{01},\beta_{11}$. If $\hat{u}$ has a double zero, we use the $x_1$-derivative of (25) at the zero as the second equation. For a germ $\hat{u}$ of degree $l$, with $l$ zeros, we obtain $l$ linear equations for $\beta_{\lambda 1}, \lambda = 0(1)l-1$, in the same fashion.

If no germ passes the test, the polynomial $\bar{p}$ has no pseudo-factorization under the specified tolerance $e$.

#### 4.1.2 Extension of a Germ into Candidate Factors

When a germ has passed one such quick test (or several ones for $s>2$), we extend it into a full pair of candidate factors $u_0, v_0$. We use the situation of (24), i.e. with $l=2$, for the explanation of the algorithm and indicate the generalization to larger $l$. Once we know $\beta_{01}$ and $\beta_{11}$ (or $\beta_{01},\ldots,\beta_{l-1,1}$ resp.), (25) yields a triangular linear system, with diagonal 1's, for the coefficients $\gamma_{\lambda 1}, \lambda = 0(1)k-l-1$, of $v_{11}$; note that the $l$ equations from the $x_1^{k-\lambda}$-terms, $\lambda = 1(1)l$, in (25) are disregarded.

Next, we consider the $x_2^2$-terms in (24):

$$\beta_{02}v_{10}(x_1)+(\beta_{01}+\beta_{11}x_1)v_{11}(x_1)+\hat{u}(x_1)v_{12}(x_1) = p_{12}(x_1). \tag{26}$$

$\beta_{02}$ (or $\beta_{02},\ldots,\beta_{l-2,2}$) are obtained upon substitution of zeros of $\hat{u}$ for $x_1$ (like previously the $\beta_{\lambda 1}$). After their computation, (26) is linear in the coefficients of $v_{12}$, with a regular matrix as for $v_{11}$ above. Again, the "superfluous" equations are dropped.

In this manner, we may compute the complete extension of the univariate germ $\hat{u}$ in the direction of one further variable (called $x_2$). For $s>2$, the same procedure applied to another variable leads to a further extension of the candidate factors. Finally, we must complete the factors $u_0, v_0$ by providing the mixed terms which do not contain the variable $x_1$ of the original germ. This can be done by using (say) $x_2$ as germ variable and extending the "germ" $u_0(0,x_2)$ in the direction of (say) $x_3$. In the generic case, we will thus obtain a complete candidate pair of factors. Naturally, all this has many parallels with classical Hensel lifting, but there are also a number of marked differences. In particular, we make no attempt to satisfy *all* relations of the overdetermined system (23) at this point.

#### 4.1.3 Checking and Correcting a Candidate Pair of Factors

By our extension procedure, we have satisfied (within round-off) all equations $q_j = 0$ in (23), with $j \in \mathbb{N}_0^s, 1 \le |j| \le k-l$, plus a number of relations corresponding to terms of a higher degree. Generally, we cannot expect that the coefficients $q_j(b_0,c_0;\bar{a})$ of the remaining terms in the *residual polynomial* $\bar{p}(x)-u_0(x)\,v_0(x)$ are small. If indeed

$$|q_j(b_0,c_0;\bar{a})| \;\le\; \varepsilon_j\,, \quad 1\le |j| \le k\,, \tag{27}$$

we have determined a *valid pseudo-factorization* of $(\bar{p},e)$ and are finished. Note that, generally, (27) cannot hold if the empirical polynomial contains some intrinsic coefficients besides the constant term 1 since this requires $|q_j| = 0$ (within round-off) for the respective $j$'s.

If (27) is not satisfied, we attempt to correct $u_0, v_0$ into a valid pseudofactor pair. We linearize *all* $q_j(b_0-\Delta b, c_0-\Delta c;\bar{a})$ of (23) about $b_0,c_0$ into $\tilde{q}_j(\Delta b,\Delta c,b_0,c_0;\bar{a})$ by dropping the $\Delta\beta_{j'}\Delta\gamma_{j''}$ terms. Then we begin to solve the *linear minimization problem*

$$\min_{\Delta b,\Delta c}\ \tilde{d}(\Delta b,\Delta c,b_0,c_0;\bar{a}) \;:=\; \min_{\Delta b,\Delta c}\ \max_{1\le|j|\le k,\ \varepsilon_j>0}\ \frac{1}{\varepsilon_j}\,|\tilde{q}_j(\Delta b,\Delta c,b_0,c_0;\bar{a})|\,, \tag{28}$$

For $j$ with $\varepsilon_j = 0$, we either pose a linear *equation condition*

$$\tilde{q}_j(\Delta b,\Delta c,b_0,c_0;\bar{a}) \;=\; 0\,, \quad \text{for } j \text{ with } \varepsilon_j = 0\,, \tag{29}$$

or we require a *fixed small* threshold $\bar{\varepsilon}_j > 0$:

$$|\tilde{q}_j(\Delta b,\Delta c,b_0,c_0;\bar{a})| \le \bar{\varepsilon}_j\,, \quad \text{for } j \text{ with } \varepsilon_j = 0\,. \tag{30}$$

During the minimization, when $\tilde{d}$ has reached a value below 1, we may stop the procedure. With $b := b_0-\Delta b, c := c_0-\Delta c$, the $q_j$ should may satisfy (27); then we have generated a valid pseudo-factorization of $(\bar{p},e)$. But due to our use of the linearized backward error, this is not guaranteed; we may have to linearize once more about the new pair $u,v$ and return to (28). Also, the value of the backward error may also stall at a value sufficiently larger than 1 and we must conclude that no pseudo-factorization can be reached from the particular germ.

### 4.2 Special Situations

We shortly discuss several situations where a modification of the approach is desirable or necessary.

#### 4.2.1 Reality and Sparsity Issues

If $p$ is real and if we are only interested in real pseudofactors, we need only consider real germs, i.e. pairs of conjugate omplex zeros must not be split between $u$ and $v$. This leads to a welcome reduction of cases in phase (i); cf. the example in section 5.2.

*Vanishing* coefficients $\bar{\alpha}_j x^j, |j| \le k$, in $\bar{p}$ will normally (but not necessarily) be associated with $\varepsilon_j = 0$ and lead to restriction (29) or (30) for (28). These will rarely *enforce* the vanishing of certain coefficients $\beta_{j'}$ and/or $\gamma_{j''}$ except if the sparsity is very systematic. Even then, one has to be careful: In the univariate case, e.g., an even polynomial has factorizations which are not even, which naturally extends to multivariate factorizations.

When the sparsity severely reduces the Newton polytope of $p$, this may imply a corresponding restriction of the Newton polytopes of potential factors. E.g., consider an $s$-variate polynomial $p$ which is of degree $k'$ *in each variable*:

$$p(x_1, \dots, x_s) \;=\; \sum_{j_1,\dots,j_s=0}^{k'} \alpha_j\, x_1^{j_1} \dots x_s^{j_s}\,.$$

Here, one will look for (pseudo)factors $u, v$ of degrees $l' < k'$ and $k' - l'$ resp. *in each variable* and modify the algorithmic approach accordingly. In his survey paper [7], E. Kaltofen poses the systematic consideration of sparsity in exact multivariate factorization as an open problem; it is as important an issue for pseudo-factorization.

#### 4.2.2 Near-Multiple Zeros in the $p(0.., x_\sigma, ..0)$

In section 4.1.1, we have assumed that potential multiple zeros of $p_{10}(x_1) = p(x_1, 0, \dots, 0)$ are not split between the germ $\hat{u}$ and the associated $v_{10}$; but this may be necessary for the pseudo-factorization. From (25), we see that the feasibility of such a germ is easily tested: Iff the multiple zero $\xi$ of $p_{10}$ is also a zero of $p_{11}$ the splitting is feasible. It necessitates a change in the determination of the $\beta_{\lambda 1}$ which we do not discuss here. Numerically, one will also expect difficulties if a *cluster* of zeros of $p_{10}$ is split between $\hat{u}$ and $v_{10}$. Obviously, this should only be considered if $p_{11}$ is very small in the cluster, with an adaptation of the procedure from the multiple zero case.

If one may select a variable $x_\sigma$ where no close zeros occur in the univariate polynomial $p_{\sigma 0}$, the difficulty can be circumvented.

#### 4.2.3 Too many Exact Data

It is clear from the general setting that we must have sufficiently many coefficients $\alpha_j$ with $\varepsilon_j > 0$ for pseudo-factorization to make sense. E.g., if the equality conditions (29) arising from the *exact* $\alpha_j$ with $\varepsilon_j = 0$ form an overdetermined system by themselves, there can generally exist a solution only if there is an exact factorization. This is not the situation in which we have been interested in this paper.

## 5. IMPLEMENTATION, EXAMPLES

### 5.1 Implementation Aspects

This is a report about work-in-progress. At submission time, our implementation consists of 5 separate Maple routines for the various tasks; no streamlining of data handling and processing has yet been attempted. Before turning this into a true floating-point procedure, with the associated further speed-up, we must design an automatic handling of the special cases which may arise. At the conference, we hope to be able to report about this in more detail.

### 5.2 A nontrivial Example

We consider a polynomial $p \in \mathbb{C}[x, y, z]$ of total degree 5:

$$\begin{aligned}
&p(x,y,z) := 1 + .75x + .42y - 1.00z - 5.50x^2 - 1.42xy\\
&-1.58y^2 - 6.63xz - .79yz - .75z^2 + 4.75x^3 + 3.75x^2y\\
&+2.25xy^2 + .58y^3 - 1.50x^2z - .50xyz + 2.50y^2z + 1.75xz^2\\
&+2.08yz^2 + 2.00z^3 + 2.00x^4 - 2.00x^3y - 1.00xy^3 + 9.50x^3z\\
&-3.00x^2yz + 3.00xy^2z - 1.00y^3z - 3.00x^2z^2 - 1.00xyz^2\\
&-1.00y^2z^2 + 3.00xz^3 - 1.00yz^3 - 1.00z^4\\
&-3.00x^5 - 1.00x^3y^2 - 1.00x^3z^2
\end{aligned}$$

which has been obtained from a factorizable polynomial $p_{\text{rat}} \in \mathbb{Q}[x, y, z]$ by rounding to 2 decimal digits so that we assume $\varepsilon_j = .005$ for the *non-vanishing* coefficients. There are only 35 non-vanishing coefficients in $p$ (of potentially 55); if we want to preserve this sparsity, we must satisfy 20 equality conditions (at least approximately). All computations have been performed as floating-point computations in `Maple`, with `digits=10`.

(i) **Quick rejection of unsuitable germs** (section 4.1.1):

Since $p(0, y, 0)$ is only of degree 3 and $p(0, 0, z)$ of degree 4, there would have been fewer germs to check with $y$ or $z$ as germ variable. But the remaining work is not affected; so we have kept the association $x_1 \leftrightarrow x$, $x_2 \leftrightarrow y$, $x_\sigma \leftrightarrow z$. However, we note that $p(0, y, 0)$ has 3 *real* zeros; this implies that all potential candidate pairs must be real because there are only linear operations in the extension phase and $p$ is real.

$p(x, 0, 0)$ has zeros (rounded) $-1.40129, -.33333,\ 1.0000,$ $.70065 \pm .47193i$, which yields 3 germs for a real linear factor and 4 germs for a real quadratic factor. With (24) and substitution into $p(0, y, 0)$, the linear germs are immediately found unsuitable (backward errors of O(100)). For a quick testing of the quadratic germs, we form the 3 possible target values for $\beta_{010}$ from $p(0, y, 0)$ (cf. Observation in section 5.1.1): -1.19803, 1.00000, 1.03803. The $\beta_{010}$ obtained from (25) are -.99597, .49924, 1.00242. With our tolerances, this leaves only the quadratic germ $\hat{u} = 1{+}2.00000\,x{-}3.00000\,x^2$ (from the zeros -.33333, 1.0000 of $p(x, 0, 0)$) for further consideration.

(ii) **Extension of the feasible germ** (section 4.1.2) :

The triangular linear systems for the extension of $\hat{u}$ w.r.t. $y$ and $z$ resp. are of sizes 3 or less! Rounded to 4 decimal digits, we obtain

$$\begin{aligned}
&u_0(x,y,z) = \ 1 + 2.0000x + 1.0024y + .4992z - 3.0000x^2\\
&-.0024xy - .9960y^2 - .0059xz + .0017yz - .9975z^2,
\end{aligned}$$

$$\begin{aligned}
&v_0(x,y,z) = \ 1 - 1.2500x - .5824y - 1.4992z + 1.0003xy\\
&-.0002y^2 - 3.0026xz + 1.0020yz + .9960z^2 + 1.0000x^3\\
&-.0008x^2y + .0013xy^2 + .0001y^3 - .0020x^2z + .0006xyz\\
&+.0035y^2z + .0008xz^2 + .0030yz^2 + .0073z^3\,.
\end{aligned}$$

(iii) **Checking and correction of $u_0, v_0$** (section 4.1.3) :

27 of the 55 equations in the overdetermined system (23) have not been used in the determination of $u_0, v_0$ and a good

number of the associated residuals are too large. Therefore, we proceed to the correction phase:

The objective function $\tilde{d}$ of the linear minimization problem (28) is the max over the moduli of 35 linear expressions in 28 correction coefficients $\Delta\beta_j, \Delta\gamma_j$, subject to 20 linear inequalities (30). It has turned out that the use of (30) in place of (29) enhances the stability of the minimization considerably. From a standard simplex solver, we obtain, rounded to 4 digits,

$$u(x,y,z) = \ 1 + 2.0007x + 1.0007y + .4991z \\ -2.9994x^2 - .9989y^2 + .0001yz - .9989z^2 \,,$$

$$v(x,y,z) = \ 1 - 1.2510x - .5827y - 1.5011z + .9997xy \\ -3.0012xz + 1.0009yz + .0002x^2 + .9994z^2 + .9995x^3,$$

with a *nonlinear* backward error $\approx .8$; the sparsity in $p$ has been reproduced within $4 \cdot 10^{-5}$. The total *computation time* (neglecting the organizational overhead in our preliminary code) has been almost negligible.

Note that these $u$ and $v$ are nearly *sparse*: 2 coefficients are missing in $u$ and 10 in $v$; they had values $\mathrm{O}(10^{-5})$ or below after the correction. With clever reasoning, this sparsity should have been *predictable*; its a-priori consideration would have reduced the computation substantially; cf. section 4.2.1.

## 6. CONCLUSIONS

We have explained and demonstrated an algorithm for the detection and computational determination of a pseudo-factorization of a multivariate polynomial $p$, i.e. of an exact factorization over $\mathbb{C}$ of some polynomial in a specified coefficientwise neighborhood of $p$. Floating-point arithmetic, with a sufficiently high precision relative to the tolerances of $p$, is to be used throughout the algorithm. We do not attempt to solve a global minimization problem with its large number of local minima; instead, we determine potential approximate candidate factors in a manner analogous to Hensel lifting but *restricted to a subsystem* of the necessary conditions. If there are any candidates, we correct them through *linearized local minimization*, with a *coefficientwise max-norm metric*. Within round-off, we can readily restrict the admissible neighborhood of $p$ to polynomials with the same sparsity pattern and/or to polynomials which agree with $p$ in specified coefficients exactly, as long as sufficiently many coefficients with a tolerance remain for adapting the factorization.

The computational complexity must increase exponentially with the total degree $k$, due to the increase in the number of factors of a univariate polynomial in $\mathbb{C}[x]$; but we eliminate each infeasible factor with a minimal computational effort. The linear algebra is again dominated by the potentially large number of coefficients, the work per coefficient is at most quadratic in the degree of the factor. The local minimization has to deal with $\binom{k+s}{s}$ linear inequalities/equations; but the minimum need only be approximated not reached. In any case, our approach appears to permit an economic treatment of more variables and higher degrees than is the state of the art ([8]). A streamlined implementation – which is not yet available – will establish the actual efficiency of the approach.

## 7. REFERENCES


[1] F. Chaitin-Chatelin, V. Frayssé: Lectures on Finite Precision Computations, Software - Environments - Tools, SIAM, Philadelphia, 1996

[2] R.M. Corless, P.M. Gianni, B.M. Trager, S.M. Watt: The singular value decomposition for polynomial systems, in: Proceed. ISSAC'95 (Ed. A.H.M.Levelt) (1995) 195-207

[3] A. Galligo, S. Watt: A Numerical Absolute Primality Test for Bivariate Polynomials, in: Proceed. ISSAC'97 (Ed. W.W.Küchlin) (1997) 217-224

[4] K.O. Geddes, S.R. Czapor, G. Labahn: Algorithms for Computer Algebra, Kluwer Academic Publ., 1992

[5] M. Giusti, É. Schost: Solving some Overdetermined Polynomial Systems, in: Proceed. ISSAC'99 (Ed. S.Dooley) (1999) 1-8

[6] M.A. Hitz, E. Kaltofen, Y.N. Lakshman: Efficient Algorithms for Computing the Nearest Polynomial with a Real Root and Related Problems, in: Proceed. ISSAC'99 (Ed. S.Dooley) (1999) 205-212

[7] E. Kaltofen: Polynomial Factorization 1987-1991, in: Proceed. LATIN'92 (Ed. I.Simon), Lect.Notes Comp.Sci. **583** (1992) 294-313

[8] E. Kaltofen: Remark in a lecture, Workshop on Symbolic and Numerical Scientific Computation (SNSC'99), August 1999.

[9] N.K. Karmarkar, Y.N. Lakshman: On Approximate GCDs of Univariate Polynomials, J. Symb.Comp. **26** (1986) 653-666

[10] T. Sasaki, M. Suzuki, M. Kolar, M. Sasaki: Approximate Factorization of Multivariate Polynomials and Absolute Irreducibility Testing, Japan J.Industr.Appl. Math. **8** (1991), 357-375

[11] T. Sasaki, T. Saito, T. Hilano: Analysis of Approximate Factorization Algorithm I, Japan J.Industr.Appl.Math. **9** (1992), 351-368

[12] T. Sasaki, F. Kako: Approximate Factorization of Multivariate Polynomials – Improvements of Naive Algorithm, unpublished communication

[13] T. Sasaki, S. Yamaguchi: An Analysis of Cancellation Error in Multivariate Hensel Construction with Floating-Point Number Arithmetic, in: Proceed. ISSAC'98 (Ed. O.Gloor) (1998) 1-8

[14] H.J. Stetter: Polynomials with Coefficients of Limited Accuracy, in: Computer Algebra in Scientific Computing (Eds. V.G.Ganzha, E.W.Mayr, E.V.Vorozhtsov), Springer, 1999, 409-430

[15] R. Zippel: Effective Polynomial Computation, Kluwer Academic Publ., 1993