



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)



Journal of Symbolic Computation 42 (2007) 693–716

---

Journal of  
Symbolic  
Computation

---

[www.elsevier.com/locate/jsc](http://www.elsevier.com/locate/jsc)

# On approximate triangular decompositions in dimension zero<sup>☆</sup>

Marc Moreno Maza<sup>a,\*</sup>, Greg Reid<sup>b</sup>, Robin Scott<sup>b</sup>, Wenyuan Wu<sup>b</sup>

<sup>a</sup> Computer Science Department, University of Western Ontario, London, ON N6A 5B7, Canada

<sup>b</sup> Applied Mathematics Department, University of Western Ontario, London, ON N6A 5B7, Canada

Received 26 November 2005; accepted 20 December 2006

Available online 12 January 2007

---

## Abstract

Triangular decompositions for systems of polynomial equations with  $n$  variables, with exact coefficients, are well developed theoretically and in terms of implemented algorithms in computer algebra systems. However there is much less research concerning triangular decompositions for systems with approximate coefficients.

In this paper we discuss the zero-dimensional case of systems having finitely many roots. Our methods depend on having approximations for all the roots, and these are provided by the homotopy continuation methods of Sommese, Verschelde and Wampler. We introduce approximate equiprojective decompositions for such systems, which represent a generalization of the recently developed analogous concept for exact systems. We demonstrate experimentally the favorable computational features of this new approach, and give a statistical analysis of its error.

© 2007 Elsevier Ltd. All rights reserved.

**Keywords:** Symbolic–numeric computations; Triangular decompositions; Dimension zero; Polynomial system solving

---

## 1. Introduction

Ritt initiated the algebraic study of differential polynomial systems through characteristic sets (Ritt, 1932). Their modern study was revitalized by the work of Wu. In Wu (1987), he

---

<sup>☆</sup> This work is supported by NSERC, MITACS, Maplesoft (Canada) and Institute of Mathematics and its Applications (University of Minnesota, USA).

\* Corresponding author.

E-mail addresses: [moreno@orcca.on.ca](mailto:moreno@orcca.on.ca) (M. Moreno Maza), [reid@uwo.ca](mailto:reid@uwo.ca) (G. Reid), [rscott2@uwo.ca](mailto:rscott2@uwo.ca) (R. Scott), [wwu25@uwo.ca](mailto:wwu25@uwo.ca) (W. Wu).

adapted the work of Ritt for solving algebraic systems: he showed that the zero set of such a system could be decomposed as finitely many characteristic sets, leading to the notion of a triangular decomposition of an algebraic variety. Considerable developments have followed from many authors; among them: Aubry et al. (1999), Chou (1988), Dahan et al. (2005), Gao and Luo (2004), Kalkbrener (1993), Lazard (1992), Moreno Maza (2000), Schost (2003), Wang (2000), and others. These works have led to efficient algorithms for triangular decomposition of an algebraic variety given by an exact input polynomial system.

Often, in applications we are interested in producing a useful triangular form where some of the variables are functions of others. Such systems frequently have approximate coefficients that are inferred from experimental data. This means that the stability, or sensitivity to coefficient changes, of such triangular decompositions is a concern. While considerable progress in both theoretical and algorithmic aspects has been made for exact input polynomial systems, much less is known about generalizations of these methods to input systems which are approximate. However, in recent years, motivated by many realistic problems, some related work has been done, for example: on numerical Gröbner bases by Stetter (2004) and the study of the approximate radical of zero-dimension ideals by Janovitz-Freireich et al. (2006).

In this paper, we present some initial results in this direction, for the case of an algebraic variety  $V$  over  $\mathbb{C}$ . We rely on the methods of Sommese, Verschelde, and Wampler (Sommese and Verschelde, 2000; Verschelde, 1999; Leykin and Verschelde, 2004; Sommese et al., 2001) which use homotopy continuation, to determine so-called generic points on the components of the numerical decomposition of  $V$ . We are interested in the set  $V_0$  of the isolated points of  $V$  (the zero-dimensional case). Each point of  $V_0$ , and more generally every irreducible component of  $V$ , is trivially a triangular set, although not generally rationally constructible from rational input. This is in contrast to the usual forms of exact triangular decomposition, which are modeled on equi-dimensional decomposition over  $\mathbb{Q}$  rather than irreducible decomposition over  $\mathbb{C}$ .

Following Dahan et al. (2004, 2005), we consider the equiprojective decomposition of  $V_0$ . Then, we use the interpolation formulas of Dahan and Schost (2004) for computing an approximate triangular set for each equiprojective component of  $V_0$ , leading to an approximate triangular decomposition of  $V_0$  in Section 3.

We provide a stability analysis of the interpolation formulas of Dahan and Schost in Section 4. One of our main tools is Lindeberg's theorem (Shiryaev, 1995) that is described in the Appendix. In Sections 5 and 6, we report on experiments that illustrate the efficiency of our approach and support the accuracy of our stability analysis.

In Moreno Maza et al. (2007), we study the simplest class of positive dimensional systems: linear homogeneous systems. Our aim in that article is to explore local structure of non-linear problems with linearized approximate triangular decompositions. The combination of the two approaches allows us to form an accessible bridge to the study of the fully non-linear case which we will describe in a forthcoming paper.

## 2. Triangular decompositions

A triangular decomposition of a zero-dimensional algebraic variety  $V$  is a family of polynomial sets, called triangular sets, that describe symbolically the points of  $V$  (Lazard, 1992). Triangular decompositions extend to algebraic varieties of arbitrary dimension; see for instance Kalkbrener (1993), Moreno Maza (2000). In Dahan and Schost (2004) it is shown that the height of a coefficient in a triangular set  $T$  can be bounded by the height of the variety represented by  $T$ . Combined with the notion of *equiprojective decomposition* introduced in Dahan et al. (2004),

this motivated the work of Dahan et al. (2005), in which the authors obtained a very efficient method for computing triangular decompositions of zero-dimensional varieties over  $\mathbb{Q}$  given by an input polynomial system with exact coefficients.

On top of these good computational properties, triangular sets and triangular decompositions have natural geometrical interpretations. In Section 3, we will rely on these properties to introduce a notion of an *approximate triangular decomposition* of a zero-dimensional variety given by approximate coordinates of its points. In the present section, we recall some results for triangular decompositions in the exact case and refer the reader to Dahan and Schost (2004), Dahan et al. (2004, 2005) for more details. For the reader's convenience, we sketch the proof of Propositions 4 and 5, which play a central role in this paper. See Dahan and Schost (2004) for their complete proofs.

Let  $\mathbb{K}$  be a perfect field, let  $\mathbb{L}$  be an algebraic closure of  $\mathbb{K}$  and let  $X_1 \prec \dots \prec X_n$  be  $n \geq 1$  ordered variables.

**Definition 1.** A set  $T = \{T_1, \dots, T_n\}$  of  $n$  polynomials in  $\mathbb{K}[X_1, \dots, X_n]$  is a *triangular set* if the ideal  $\langle T \rangle$  generated by  $T$  is radical and if for all  $1 \leq i \leq n$  the polynomial  $T_i$  is not constant, the greatest variable occurring in  $T_i$  is  $X_i$ , and its leading coefficient w.r.t.  $X_i$  is invertible modulo the ideal  $\langle T_1, \dots, T_{i-1} \rangle$ . The triangular set  $T$  is *normalized* if for all  $1 \leq i \leq n$  the leading coefficient of  $T_i$  w.r.t.  $X_i$  is 1.

Clearly, a triangular set generates a zero-dimensional ideal and a normalized triangular set is a reduced lexicographical Gröbner basis. In Lazard (1992), it is shown that every maximal ideal of  $\mathbb{K}[X_1, \dots, X_n]$  can be generated by a triangular set. Hence, a natural question is how to characterize the zero-dimensional varieties over  $\mathbb{K}$ , that can be generated by a triangular set. The answer is given by Aubry and Valibouze (2000). We report on it here by means of Definition 2 and Theorem 3, after introducing some notation.

Let  $i$  and  $j$  be integers such that  $1 \leq i \leq j \leq n$ . We denote by  $A^i(\mathbb{L})$  the affine space of dimension  $i$  over  $\mathbb{L}$ . For  $V \subseteq A^n(\mathbb{L})$  we denote by  $\mathcal{I}(V)$  the ideal of  $\mathbb{K}[X_1, \dots, X_n]$  composed by the polynomials which vanish on  $V$ . For  $F \subseteq \mathbb{K}[X_1, \dots, X_n]$  we denote by  $V(F)$  the set of the points of  $A^n(\mathbb{L})$  where every element of  $F$  vanishes. Finally, we denote by  $\pi_i^j$  the natural projection map from  $A^j(\mathbb{L})$  to  $A^i(\mathbb{L})$ , which sends  $(X_1, \dots, X_j)$  to  $(X_1, \dots, X_i)$ .

**Definition 2.** A zero-dimensional variety  $V \subseteq A^j(\mathbb{L})$  over  $\mathbb{K}$  is said to be:

- (1) *equiprojectable on  $V_i = \pi_i^j(V)$* , its projection onto  $A^i(\mathbb{L})$ , if there exists an integer  $c$  such that for every  $M \in V_i$  the cardinality of  $(\pi_i^j)^{-1}(M) \cap V$  is  $c$ ;
- (2) *equiprojectable* if  $V$  is equiprojectable on  $V_1, \dots, V_{j-1}$ .

**Theorem 3.** A zero-dimensional variety  $V \subseteq A^j(\mathbb{L})$  over  $\mathbb{K}$  is equiprojectable if and only if there exists a triangular set  $T$  of  $\mathbb{K}[X_1, \dots, X_j]$  such that  $T$  generates  $\mathcal{I}(V)$ .

Given an equiprojectable variety  $V \subseteq A^n(\mathbb{L})$  the normalized triangular set  $T$  generating  $\mathcal{I}(V)$  can be constructed as follows from the coordinates of the points of  $V$  (see Dahan and Schost (2004) for details). Let  $\mathbf{K}$  be a field such that  $\mathbb{K} \subseteq \mathbf{K} \subseteq \mathbb{L}$  and such that every point of  $V$  has its coordinates in  $\mathbf{K}$ . We define  $V_i = \pi_i^n(V)$ . Let  $1 \leq \ell < n$ . Following Dahan and Schost (2004), we describe how to interpolate  $T_{\ell+1}$  from the coordinates (in  $\mathbf{K}$ ) of the points of  $V_{\ell+1}$ .

Let  $\alpha = (\alpha_1, \dots, \alpha_\ell) \in V_\ell$ . Define

$$\begin{aligned} V_\alpha^1 &= \{\beta = (\beta_1, \dots, \beta_\ell, \beta_{\ell+1}) \in V_{\ell+1} \mid \beta_1 \neq \alpha_1\}, \\ V_\alpha^2 &= \{\beta = (\alpha_1, \beta_2, \dots, \beta_\ell, \beta_{\ell+1}) \in V_{\ell+1} \mid \beta_2 \neq \alpha_2\}, \\ V_\alpha^3 &= \{\beta = (\alpha_1, \alpha_2, \beta_3, \dots, \beta_\ell, \beta_{\ell+1}) \in V_{\ell+1} \mid \beta_3 \neq \alpha_3\}, \\ &\dots \dots \dots \\ V_\alpha^\ell &= \{\beta = (\alpha_1, \dots, \alpha_{\ell-1}, \beta_\ell, \beta_{\ell+1}) \in V_{\ell+1} \mid \beta_\ell \neq \alpha_\ell\}, \\ V_\alpha^{\ell+1} &= \{\beta = (\alpha_1, \dots, \alpha_\ell, \beta_{\ell+1}) \in V_{\ell+1}\}. \end{aligned} \tag{1}$$

The sets  $V_\alpha^1, V_\alpha^2, V_\alpha^3, \dots, V_\alpha^\ell, V_\alpha^{\ell+1}$  partition  $V_{\ell+1}$ . We consider also the projections

$$\begin{aligned} v_\alpha^1 &= \pi_1^{\ell+1}(V_\alpha^1) = \{(\beta_1) \in V_1 \mid \beta_1 \neq \alpha_1\}, \\ v_\alpha^2 &= \pi_2^{\ell+1}(V_\alpha^2) = \{(\alpha_1, \beta_2) \in V_2 \mid \beta_2 \neq \alpha_2\}, \\ v_\alpha^3 &= \pi_3^{\ell+1}(V_\alpha^3) = \{(\alpha_1, \alpha_2, \beta_3) \in V_3 \mid \beta_3 \neq \alpha_3\}, \\ &\dots \dots \dots \\ v_\alpha^\ell &= \pi_\ell^{\ell+1}(V_\alpha^\ell) = \{(\alpha_1, \dots, \alpha_{\ell-1}, \beta_\ell) \in V_\ell \mid \beta_\ell \neq \alpha_\ell\}. \end{aligned} \tag{2}$$

For  $1 \leq i \leq \ell + 1$ , we define

$$T_{\alpha,i} = T_i(\alpha_1, \dots, \alpha_{i-1}, X_i) \quad \text{and} \quad e_{\alpha,i} = \prod_{\beta \in v_\alpha^i} (X_i - \beta_i). \tag{3}$$

Observe that for  $1 \leq i \leq \ell + 1$  we have  $T_{\alpha,i} \in \mathbf{K}[X_i]$  and  $e_{\alpha,i} \in \mathbf{K}[X_i]$ . Finally, we define

$$E_\alpha = \prod_{1 \leq i \leq \ell} e_{\alpha,i} \tag{4}$$

and note that  $E_\alpha \in \mathbf{K}[X_1, \dots, X_\ell]$  holds.

**Proposition 4.** For  $1 \leq i \leq \ell$  we have

$$T_{\alpha,i} = \prod_{(\alpha_1, \dots, \alpha_{i-1}, \beta_i) \in V_i} (X_i - \beta_i) = e_{\alpha,i} (X_i - \alpha_i), \tag{5}$$

$$T_{\alpha,\ell+1} = \prod_{\beta \in V_\alpha^{\ell+1}} (X_{\ell+1} - \beta_{\ell+1}), \tag{6}$$

$$T_{\ell+1} = \sum_{\alpha \in V_\ell} \frac{E_\alpha T_{\alpha,\ell+1}}{E_\alpha(\alpha)}. \tag{7}$$

**Proof.** Relations (5) and (6) follow easily from (1)–(3). In order to prove (7) we observe that

$$(\forall \beta \in V_\ell) E_\alpha(\beta) = 0 \iff \beta \neq \alpha. \tag{8}$$

Indeed, for  $1 \leq i \leq \ell$ , we have  $e_{\alpha,i}(\alpha) \neq 0$  leading to  $E_\alpha(\alpha) \neq 0$ . Now let  $\beta \in V_\ell$  with  $\beta \neq \alpha$ . Then, there exists  $i \leq \ell$  such that

$$(\pi_i^\ell)^{-1}(\beta) \in v_\alpha^i.$$

Hence, for this index  $i$  we have  $e_{\alpha,i}(\beta) = 0$ , which proves (8). From there, establishing (7) is routine.  $\square$

In Dahan and Schost (2004), another triangular set  $N$  is obtained from the coordinates of the points of  $V$ ; see Proposition 5. The authors show that it has much smaller coefficients than the normalized triangular set given by the formulas of Proposition 4. We will be generalizing this second triangular set to the approximate case.

**Proposition 5 (Interpolation Formulas).** *Let  $D_1 = 1$  and  $\tau_1 = N_1 = T_1$ . For  $2 \leq \ell \leq n$ , define*

$$D_\ell = \prod_{1 \leq i \leq \ell-1} \frac{\partial T_i}{\partial X_i} \pmod{\langle T_1, \dots, T_{\ell-1} \rangle} \quad (9)$$

and

$$N_\ell = D_\ell T_\ell \pmod{\langle T_1, \dots, T_{\ell-1} \rangle}. \quad (10)$$

Then, for  $1 \leq i \leq \ell$  we have

$$N_{\ell+1} = \sum_{\alpha \in V_\ell} E_\alpha T_{\alpha, \ell+1}. \quad (11)$$

**Proof.** Indeed, for  $1 \leq i \leq \ell$ , we have

$$T_{\alpha, i} = e_{\alpha, i}(X_i - \alpha_i) \in \mathbf{K}[X_i]$$

leading to

$$\begin{aligned} \frac{\partial T}{\partial X_i}(\alpha) &= T'_{\alpha, i}(\alpha) \\ &= e'_{\alpha, i}(\alpha)(\alpha_i - \alpha_i) + e_{\alpha, i}(\alpha) \\ &= e_{\alpha, i}(\alpha). \end{aligned}$$

By definition, we have

$$N_{\ell+1} = \left( \prod_{1 \leq i \leq \ell} \frac{\partial T}{\partial X_i} \right) T_{\ell+1} \pmod{\langle T_1, \dots, T_\ell \rangle}.$$

Hence, we have

$$\begin{aligned} N_{\ell+1}(\alpha) &= \left( \prod_{1 \leq i \leq \ell} \frac{\partial T}{\partial X_i}(\alpha) \right) T_{\ell+1}(\alpha) \\ &= \left( \prod_{1 \leq i \leq \ell} e_{\alpha, i}(\alpha) \right) T_{\ell+1}(\alpha) \\ &= E_\alpha(\alpha) T_{\ell+1}(\alpha) \end{aligned}$$

where  $T_{\ell+1}(\alpha) = T_{\alpha, \ell+1}$  holds. Finally we obtain

$$\begin{aligned} N_{\ell+1} &= \sum_{\alpha \in V_\ell} \frac{E_\alpha N_{\ell+1}(\alpha)}{E_\alpha(\alpha)} \\ &= \sum_{\alpha \in V_\ell} E_\alpha T_{\ell+1}(\alpha). \quad \square \end{aligned}$$

Clearly, not all zero-dimensional varieties over  $\mathbb{Q}$  are equiprojective. Consider, for example, with  $n = 2$  the variety consisting of the three points  $A, B, C$  with respective coordinates  $(1, 0)$ ,  $(0, 0)$  and  $(0, 1)$ . However, we do have the following result; see for instance Lazard (1992).

**Proposition 6.** *For every zero-dimensional radical ideal  $\mathcal{I}$  of  $\mathbb{K}[X_1, \dots, X_n]$  there exists finitely many triangular sets  $T^1, \dots, T^e$  such that  $\mathcal{I}$  is the intersection of the ideals  $\langle T^1 \rangle, \dots, \langle T^e \rangle$ . If, in addition, the ideals  $\langle T^1 \rangle, \dots, \langle T^e \rangle$  are pairwise relatively prime, then the set  $\{T^1, \dots, T^e\}$  is called a triangular decomposition of the ideal  $\mathcal{I}$ .*

Triangular decompositions of algebraic varieties (with arbitrary dimension) are discussed in depth in Moreno Maza (2000) together with an algorithm for computing them, which is implemented in Lemaire et al. (2005). Observe that a radical ideal may admit several triangular decompositions. For instance, there are four different triangular decompositions for the ideal  $\mathcal{I}(\{A, B, C\})$ . Choosing a canonical triangular decomposition for the radical  $\mathcal{I}$  with the variable ordering  $X_1 \prec \dots \prec X_n$  is achieved by the following combinatorial construction. We refer the reader to Dahan et al. (2005) for a more formal definition.

**Definition 7.** Consider a zero-dimensional variety  $V$  and denote by  $\pi = \pi_{n-1}^n$  the projection which removes the last coordinate. With a point  $x$  in  $V$ , we associate  $N(x) = \#\pi^{-1}(\pi(x))$ , that is, the number of points lying in the same  $\pi$ -fiber as  $x$ . Then, we split  $V$  into the disjoint union  $V_1 \cup \dots \cup V_d$ , where for all  $i = 1, \dots, d$ ,  $V_i$  equals  $N^{-1}(i)$ , that is, the set of points  $x \in V$  which have  $N(x) = i$ . This splitting process is applied recursively to all varieties  $V_1, \dots, V_d$ , taking into account the fibers of the successive projections  $\pi_i^n$ , for  $i = n-1, \dots, 1$ . In the end, we obtain a family of pairwise disjoint, equiprojective varieties, whose reunion equals  $V$ ; they form the *equiprojective decomposition* of  $V$ .

### 3. Approximate equiprojective decomposition in dimension zero

In this section, we consider a zero-dimensional variety  $V \subseteq A^n(\mathbb{C})$  over  $\mathbb{Q}$ . Each point of  $V$  is given by approximate coordinates in a sense that we make precise in Definition 8. We aim at defining and computing an *approximate triangular decomposition* of  $V$ . To do so, we extend the construction given by Definition 7 and introduce a notion of an *approximate equiprojective decomposition* of  $V$  in Definition 14. Then, with each approximate equiprojective component, we associate an approximate triangular set, leading to Definition 15 of an *approximate triangular decomposition* of  $V$ .

Therefore, an approximate triangular decomposition of  $V$  is obtained by interpolating the points of  $V$  given by approximate coordinates. We provide stability analysis for this interpolation in Section 4. Moreover, we report on experiments that illustrate the accuracy of our stability analysis in Sections 5 and 6.

**Definition 8.** Let  $\epsilon > 0$  and  $r \geq 0$  be real numbers. Let  $\bar{x} = (\bar{x}_1, \dots, \bar{x}_n)$  be a point of  $V$  and let  $x = (x_1, \dots, x_n) \in A^n(\mathbb{C})$  with  $x \neq 0$ . We say that  $(x, r)$  is an *approximate point* for  $\bar{x}$  with tolerance  $\epsilon$ , denoted by  $\bar{x} \simeq_\epsilon (x, r)$ , if the following conditions hold for all  $1 \leq i \leq n$ :

- (i)  $|\bar{x}_i - x_i| \leq r$ ,
- (ii)  $r \leq \epsilon|x|$ ,

where  $|x| = \max(|x_1|, \dots, |x_n|)$ .

With the notation of [Definition 8](#) let  $(x, r)$  be an approximate point for  $\bar{x}$  with tolerance  $\epsilon$ . Let  $1 \leq i \leq n$  be fixed. If  $\bar{x}_i$  and  $x_i$  are complex numbers and  $\bar{x}_i \neq 0$  then a frequently used measure of the number of correct significant decimal digits in the approximate coordinate  $x_i$  is the *logarithm of the relative error*  $\text{lre}(x_i, \bar{x}_i)$  given by

$$\text{lre}(x_i, \bar{x}_i) = -\log_{10} \frac{|\bar{x}_i - x_i|}{|\bar{x}_i|}. \quad (12)$$

Properties (i) and (ii) of [Definition 8](#) lead to

$$\text{lre}(x_i, \bar{x}_i) \geq -\log_{10} \epsilon - \log_{10} \frac{|x_i|}{|\bar{x}_i|}. \quad (13)$$

In practice, one requires  $\epsilon < 1$  and thus Formula (13) gives a good measure of the approximation of coordinate  $\bar{x}_i$  by means of coordinate  $x_i$ . Similarly, Formula (14) below gives a good measure of the approximation of point  $\bar{x}$  by means of point  $x$ , for  $x \neq 0$ :

$$\text{lre}(x, \bar{x}) = -\log_{10} \frac{|\bar{x} - x|}{|x|}. \quad (14)$$

As we shall see now, another good measure of this approximation is

$$\text{lb}(\bar{x}, x) = -\log_{10} \frac{|\bar{x} - x|}{|x|}. \quad (15)$$

Indeed, one can easily check that the following holds:

$$\left| \log_{10} \frac{|\bar{x} - x|}{|x|} - \log_{10} \frac{|\bar{x} - x|}{|\bar{x}|} \right| = \left| \log_{10} \frac{|\bar{x}|}{|x|} \right|. \quad (16)$$

Moreover, we claim that when  $\epsilon$  is close to zero,

$$\left| \log_{10} \frac{|\bar{x}|}{|x|} \right| \approx \epsilon. \quad (17)$$

Thus,  $\text{lre}(x, \bar{x})$  and  $\text{lb}(\bar{x}, x)$  are very close when  $\epsilon$  is very small. To prove our claim, we start from

$$||\bar{x}| - |x|| \leq |\bar{x} - x| \leq \epsilon |x|, \quad (18)$$

which holds by assumption (points (i) and (ii) of [Definition 8](#)). We deduce

$$\left| \frac{|\bar{x}|}{|x|} - 1 \right| \leq \epsilon. \quad (19)$$

Since  $\epsilon$  is meant to be very small, using  $\log_{10}(1 - \epsilon) \approx -\epsilon$  and  $\log_{10}(1 + \epsilon) \approx \epsilon$ , we finally obtain Formula (17).

A representation (using approximate points in the sense of [Definition 8](#)) of the isolated roots of the variety  $V \subseteq A^n(\mathbb{C})$  of an input polynomial system  $F = \{F_1, \dots, F_n\} \subset \mathbb{Q}[X_1, \dots, X_n]$  can be obtained by numerical homotopy construction. In particular, we used the PHC software ([Verschelde, 1999](#)). Indeed, for each point  $\bar{x}$  of  $V$ , the corresponding solution  $x$  returned by PHC is given with the condition number of the Jacobian matrix of  $F$  at  $x$ , denoted by *cond*. The value *cond* can be used to estimate the distance between  $\bar{x}$  and  $x$  (see [Leykin and Verschelde \(2004\)](#) for details). More precisely, because we use double-precision floating-point numbers in the computation, a reasonable formula is:  $|\bar{x}_i - x_i|/|x_i| \approx \text{cond} \times 10^{-16}$  for all  $1 \leq i \leq n$

(see Table 4). Given  $\epsilon > 0$ , with this estimate, one can check whether each isolated point  $\bar{x}$  of  $V$  admits approximate points within tolerance  $\epsilon$ . Theoretically, the homotopy continuation method can obtain approximate points arbitrarily close to the exact roots for any tolerance  $\epsilon$ . So, if the multiplicity of each point is 1, a one-to-one map between approximate roots and exact ones can be computed. Note that none of the systems used in Section 6 have multiple roots (see Table 2).

**Remark 9.** The definition of approximate points of a polynomial systems is related to alpha-theoretic concepts of approximate zero (Blum et al., 1997). Although alpha theory can determine a basin in which Newton's method is guaranteed to converge, we note that our approximate zero is not necessarily in the basin of attraction of the given root. Another related concept is that of “pseudozero domains”, as introduced by Stetter to make a general study of the data to result maps in the context of the Numerical Polynomial Algebra (Stetter, 2004). In particular, we consider only local properties (especially in the stability analysis) specifically aimed at the tasks for our paper.

Let  $\epsilon > 0$ . From now on, we assume that for each point  $\bar{x} \in V$  we are given  $x \in A^n(\mathbb{C})$  and  $r > 0$ , such that  $\bar{x} \simeq_\epsilon (x, r)$  holds. Then, we denote by  $\tilde{V}$  the set of all  $(x, r)$ , and we write  $V \simeq_\epsilon \tilde{V}$ .

We now return to the construction given by Definition 7. Again let  $\pi = \pi_{n-1}^n$  be the natural projection from  $A^n(\mathbb{C})$  to  $A^{n-1}(\mathbb{C})$  which removes the last coordinate. Given two points  $\bar{x}$  and  $\bar{x}'$  of  $V$  we have to decide whether they lie in the same  $\pi$ -fiber. Since  $\bar{x}$  and  $\bar{x}'$  are given by approximate points we need the following.

**Definition 10.** Let  $i$  and  $j$  be integers such that  $1 \leq i \leq j \leq n$ . Let  $\bar{x}, \bar{y} \in \pi_j^n(V)$ . Let  $x = (x_1, \dots, x_j)$  (resp.  $y = (y_1, \dots, y_j)$ ) and  $(x, r)$  (resp.  $(y, r')$ ) be approximate coordinates of  $\bar{x}$  (resp.  $\bar{y}$ ) with tolerance  $\epsilon$ . We say that  $\bar{x}$  and  $\bar{y}$  lie approximately in the same  $\pi_i^j$ -fiber with tolerance  $\epsilon$  if for all  $1 \leq k \leq i$  we have

$$|x_k - y_k| \leq r + r'. \quad (20)$$

**Proposition 11.** With the notation of Definition 10, if the points  $\bar{x}, \bar{y} \in \pi_j^n(V)$  are in the same  $\pi_i^j$ -fiber, that is, if  $\pi_i^j(\bar{x}) = \pi_i^j(\bar{y})$ , then the points  $\bar{x}$  and  $\bar{y}$  lie approximately in the same  $\pi_i^j$ -fiber with tolerance  $\epsilon$ .

**Proof.** Since  $\bar{x}$  and  $\bar{y}$  are in the same  $\pi_i^j$ -fiber and supposing  $(x, r)$  (resp.  $(y, r')$ ) are the approximate coordinates of  $\bar{x}$  (resp.  $\bar{y}$ ) with tolerance  $\epsilon$ , then, for any  $1 \leq k \leq i$ , this leads to:

$$|x_k - y_k| = |x_k - y_k - \bar{x}_k + \bar{y}_k| \leq |\bar{x}_k - x_k| + |\bar{y}_k - y_k| \leq r + r'. \quad \square \quad (21)$$

**Remark 12.** Suppose  $1 \leq i \leq j \leq n$ . For the points of  $\pi_j^n(V)$ , the relation “lying approximately in the same  $\pi_i^j$ -fiber with tolerance  $\epsilon$ ” may not be an equivalence relation, since the transitivity axiom does not hold here. We need to exclude this situation in order to adapt the construction of Definition 7 for the points of  $V$  to approximate points of  $V$ . In theory, for exact systems, this situation may be avoided by reducing the tolerance  $\epsilon$ , and thus the radius  $r$  at each point of  $V$ . However, in practice, for some systems it is hard to obtain approximate roots when  $\epsilon$  is very small. For example, for systems possessing a cluster of points, it can be difficult to compute these roots with high precision (Leykin et al., 2005). Additionally, for input systems with limited accuracy, a tolerance beyond this limit could not be achieved. So for such systems, we would not

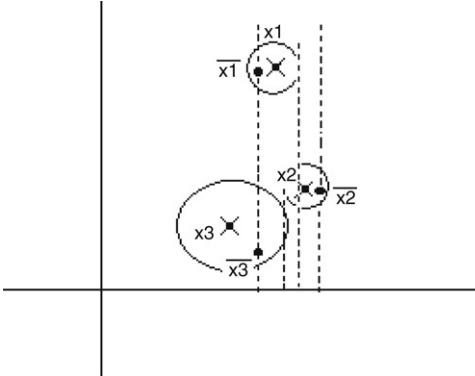


Fig. 1.  $\bar{x}_1, \bar{x}_2, \bar{x}_3$  are exact points,  $x_1, x_2, x_3$  are the approximate points respectively. Here,  $\bar{x}_1, \bar{x}_2$  lie in different fibers, but are approximately in the same fiber, and  $\tilde{V}$  satisfies the weak equivalence condition.

be able to meet the requirements of [Definition 14](#). These precautionary remarks being made, we will propose in [Definition 14](#) a notion of an *approximate equiprojective decomposition* of  $V$ , where the points of  $V$  are given by approximate points in the sense of [Definition 8](#).

For any zero-dimensional system, using some random linear coordinate change, each fiber has only one point. However, changes of coordinates will generally destroy the sparsity of the original systems. An alternative approach for avoiding unfavorable projections is to view a cluster as a perturbed multiple solution (e.g. see the recent work of [Janovitz-Freireich et al. \(2006\)](#)).

**Definition 13.** We say that  $\tilde{V}$  satisfies the *weak equivalence condition with tolerance  $\epsilon$*  if for all  $1 \leq i \leq j \leq n$ , the relation “lying approximately in the same  $\pi_i^n$ -fiber with tolerance  $\epsilon$ ” is an equivalence relation in  $\pi_j^n(V)$ . Furthermore, we say that  $\tilde{V}$  satisfies the *strong equivalence condition with tolerance  $\epsilon$*  if for every  $\bar{x}, \bar{y} \in V$  with approximate points  $(x, r), (y, r') \in \tilde{V}$ , with tolerance  $\epsilon$ , for all  $1 \leq j \leq n$  the following conditions are equivalent:

- we have  $\pi_j^n(\bar{x}) = \pi_j^n(\bar{y})$ ,
- the points  $\bar{x}$  and  $\bar{y}$  lie approximately in the same  $\pi_j^n$ -fiber.

Here we illustrate [Definition 13](#) through [Figs. 1–3](#) where we consider different  $\tilde{V}$ ’s for the same  $V$ . In [Fig. 1](#), the set  $\tilde{V}$  satisfies the weak equivalence condition; observe that  $\bar{x}_1, \bar{x}_2$  lie approximately in the same fiber, but  $\bar{x}_1$  and  $\bar{x}_2$  lie in different fibers. In [Fig. 2](#), the points  $x_1, \bar{x}_2$  and  $\bar{x}_1, x_3$  are pairs of points lying approximately in the same fiber, but  $\bar{x}_2, x_3$  do not lie approximately in the same fiber. Hence, in this case, the set  $\tilde{V}$  does not satisfy the weak equivalence condition. In [Fig. 3](#), we refine the three approximate roots until the weak equivalence condition is satisfied again (the strong equivalence condition is also satisfied); we see that  $\bar{x}_1, \bar{x}_2$  lie in different fibers.

In practice, the “exact” points of  $V$  are unknown, so we cannot determine whether the strong equivalence condition is satisfied or not. However, we can detect whether the weak equivalence condition holds or not. In our experiments reported in [Section 6](#), however, the exact points are known for each variety  $V$ , and we could decide whether or not  $\tilde{V}$  satisfies the strong equivalence condition.

If the weak equivalence condition is satisfied but the strong equivalence condition is not (e.g. see [Fig. 1](#)), then there exist two distinct points  $\bar{x}, \bar{y} \in V$ , with respective approximate points  $(x, r), (y, r')$ , and an index  $1 \leq i \leq n$  such that  $\bar{x}_i$  and  $\bar{y}_i$  are different but very close to each

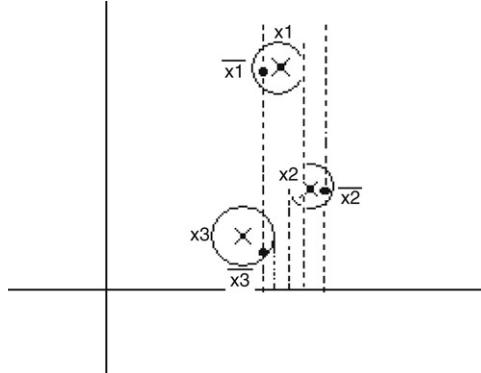


Fig. 2. Refining  $x_3$  we get a smaller radius. Here, the two pairs  $\bar{x}_1, \bar{x}_2$  and  $\bar{x}_1, \bar{x}_3$  lie approximately in the same fiber, but  $\bar{x}_2, \bar{x}_3$  do not lie approximately in the same fiber. The set  $\tilde{V}$  does not satisfy the weak equivalence condition.

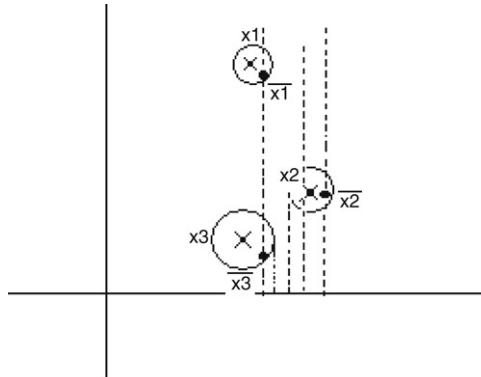


Fig. 3. Refining  $x_1$  we get the correct result. Here,  $\bar{x}_1, \bar{x}_2$  lie in different fibers and both weak and strong equivalence conditions are satisfied.

other; more precisely  $|\bar{x}_i - \bar{y}_i| < 2r + 2r'$  holds (generally the distance  $|\bar{x}_i - \bar{y}_i|$  will be less than  $10^{-13}$ ; see Table 4). Due to round-off errors in numerical computation, we cannot always avoid these rare cases.

Finally, we note that introducing the notion of “weak equivalence condition” is needed by Definition 14.

**Definition 14.** Assume that  $\tilde{V}$  satisfies the weak equivalence condition with tolerance  $\epsilon$ . Define  $\pi = \pi_{n-1}^n$ . With every point  $\bar{x}$  in  $V$ , we associate  $N(\bar{x})$  the number of points in  $V$  which lie approximately in the same  $\pi$ -fiber as  $x$  with tolerance  $\epsilon$ . For all  $i \geq 1$ , we denote by  $V_i$  the set of points  $x \in V$  satisfying  $N(x) = i$ . Then, we split  $V$  into a disjoint union  $V_1 \cup \dots \cup V_d$ , for some  $d \in \mathbb{N}$  large enough. This splitting process is applied recursively to all  $V_1, \dots, V_d$ , taking into account the fibers of the successive projections  $\pi_i^n$ , for  $i = n-1, \dots, 1$ . In the end, we obtain a family of pairwise disjoint subsets of  $V$ , whose union equals  $V$ ; they form an *approximate equiprojective decomposition* of  $V$  with tolerance  $\epsilon$ . If this approximate equiprojective decomposition of  $V$  (with tolerance  $\epsilon$ ) consists of only one subset, that is,  $V$  itself, we say that  $V$  is *equiprojective* with tolerance  $\epsilon$ ; otherwise the parts of the approximate

equiprojective decomposition of  $V$  (with tolerance  $\epsilon$ ) are called *approximate equiprojective components* of  $V$  with tolerance  $\epsilon$ .

Note that each approximate equiprojective component of  $V$  is equiprojective with tolerance  $\epsilon$ . With each approximate equiprojective component of  $V$  with tolerance  $\epsilon$  we can associate an *approximate triangular set* by means of [Definition 15](#). This leads to a notion of an *approximate triangular decomposition* for the variety  $V$ .

**Definition 15.** Assume that the zero-dimensional variety  $V$  is equiprojective with tolerance  $\epsilon$ . Then, by means of the interpolation formulas of [Proposition 5](#) one can compute a triangular set  $\{N_1, \dots, N_n\}$  called an *approximate triangular set* of  $V$  with tolerance  $\epsilon$ .

Now, assume that  $V$  is not approximately equiprojective with tolerance  $\epsilon$ . A family of approximate triangular sets of approximate equiprojective components of  $V$  (with tolerance  $\epsilon$ ) forms an *approximate triangular decomposition* of  $V$ , with tolerance  $\epsilon$ .

#### 4. Stability analysis

In this section, we explore the relation between the *relative error* on the coordinates of the approximate points of  $V$  and the *relative error* on the interpolated polynomials of the approximate triangular decomposition given by [Definition 15](#). The coefficients of a polynomial continuously depend on its roots. However, a small error in a root may result in a large error in the coefficients, motivating some of stability analysis.

For the relation between the errors mentioned above to be useful in practice, we must face the following fact: the relative error of a root cannot be computed when the exact root is unknown. In order to overcome this difficulty, for a point  $\bar{x}$  of  $V$  given by an approximate point  $(x, r)$ , we view the exact coordinates  $\bar{x} = (\bar{x}_1, \dots, \bar{x}_n)$  as a random variable which takes values in the region defined by the following: for all  $1 \leq i \leq n$

$$|x_i - \bar{x}_i| \leq r. \quad (22)$$

In this paper, we used the word *bias* instead of *relative error* in order to avoid conflicting terminology.

**Definition 16.** For  $\bar{x}, x \in \mathbb{C}$ , we define as the *bias* of  $x$  w.r.t.  $\bar{x}$  the fraction

$$\delta_x = \frac{\bar{x} - x}{x} \quad (23)$$

simply denoted by  $\delta$ , when no confusion may occur.

**Remark 17.** We would like to observe at this point that none of the results of this section require knowledge of the exact coordinates of the points of  $V$ . Hence, our results apply also in practice to the situation where  $V$  is initially given by a polynomial system with inexact coefficients rather than a polynomial system with exact coefficients. Note that the PHC software ([Verschelde, 1999; Leykin and Verschelde, 2004](#)) can process both kinds of polynomial systems.

We define now the bias for the coefficients of a polynomial. Our definition applies to univariate polynomials as well as to multivariate ones. Let  $e = (e_1, \dots, e_n) \in \mathbb{N}^n$  be an exponent vector. We denote by  $X^e$  the monomial  $X_1^{e_1} \cdots X_n^{e_n}$  of  $\mathbb{C}[X_1, \dots, X_e]$ . We write  $p = \sum_{e \in S} f_e X^e$ , a polynomial of  $\mathbb{C}[X_1, \dots, X_e]$  with (finite) support  $S$ . For every  $e \in \mathbb{N}^n$  with  $e \notin S$  we set to zero the coefficient  $f_e$ , i.e. we define  $f_e = 0$ . Hence we can simply write  $p = \sum_e f_e X^e$ .

Typically, in our stability analysis, the polynomial  $f$  of [Definition 18](#) will be a polynomial interpolating the approximate coordinates of the points of  $V$ , whereas  $\bar{f}$  will be the corresponding polynomial obtained from the exact coordinates of the points of  $V$ .

**Definition 18.** Let  $\bar{p} = \sum_e \bar{f}_e X^e$  and  $p = \sum_e f_e X^e$  be polynomials in  $\mathbb{C}[x_1, \dots, x_e]$ . For every  $e \in \mathbb{N}^n$ , the *bias of coefficient  $f_e$*  w.r.t.  $\bar{p}$  is defined by

$$\delta_e = \frac{\bar{f}_e - f_e}{f_e}. \quad (24)$$

The *bias of the polynomial  $p$*  w.r.t.  $\bar{p}$  is the bias of the coefficient of  $p$  w.r.t.  $\bar{p}$  which has the largest norm.

The interpolated polynomials given by [Proposition 5](#) are multivariate polynomials that are constructed as univariate ones over a suitable coefficient ring. Because of these formulas, we can focus on the univariate case. Let  $\bar{p} \in \mathbb{C}[X]$  be a univariate monic polynomial of degree  $b$  given by approximate values  $x_1, \dots, x_b$  of its roots with respective radii  $r_1, \dots, r_b$ :

$$p = \prod_{i=1}^{i=b} (x - x_i). \quad (25)$$

Let  $\delta_1, \dots, \delta_b$  be the respective biases of  $x_1, \dots, x_b$  such that the exact roots of  $\bar{p}$  are  $x_1 + x_1 \delta_1, \dots, x_b + x_b \delta_b$ . Hence we have

$$\bar{p} = \prod_{i=1}^{i=b} (x - x_i - x_i \delta_i). \quad (26)$$

**Notation 1.** In the remainder of this section, we assume that  $\delta_1, \dots, \delta_b$  are independent random (complex) variables, each of them with uniform distribution in a disk centered at 0 and with respective radii  $r_1/|x_1|, \dots, r_b/|x_b|$ . We define the *bias bound* and we denote it by  $\rho$ , the maximum of  $r_1/|x_1|, \dots, r_b/|x_b|$ .

In the proofs of [Propositions 19, 21](#) and [22](#), we will denote by  $O(\delta^2)$  any term in  $\delta_i \delta_j$ . When  $\rho$  is very small, we can ignore such higher order terms and keep only the linear terms.

We will consider the bias of the polynomial  $\bar{p}$  w.r.t.  $p$  as a random variable denoted by  $\gamma$ . We direct the reader to the [Appendix](#) for a brief review of the standard probability results which will be used.

There are essentially three steps in computing the interpolated polynomials of [Proposition 5](#):

- (I1) compute the univariate polynomials  $e_{\alpha,i}$ ,
- (I2) compute the multivariate polynomials  $E_\alpha$ , which are products of univariate polynomials  $e_{\alpha,i}$ ,
- (I3) compute the multivariate polynomials  $N_\ell$  which are sums of some multivariate polynomials.

For each step, we provide properties on the stability analysis of the corresponding calculations. For our study of the relation between  $\bar{p}$  and  $p$ , we need the following notation.

**Notation 2.** For  $1 \leq k \leq b$ , the  $k$ -th elementary symmetric function of  $x_1, \dots, x_b$  is given by

$$\sigma^k = \sum_{1 \leq a_1 < a_2 < \dots < a_k \leq b} x_{a_1} \cdots x_{a_k}, \quad (27)$$

and let  $\sigma^0 := 1$ . Observe that we have

$$p = \prod_{i=1}^b (x - x_i) = \sum_{k=0}^b (-1)^k \sigma^k x^{b-k}. \quad (28)$$

Let  $1 \leq j \leq b$ . We denote by  $\sigma_j^k$  the element of  $\mathbb{C}[x_1, \dots, x_n]$  obtained from  $\sigma^k$  by specializing  $x_j$  to 0, that is  $\sigma_j^k = \sigma^k|_{x_j=0}$ . Let  $l_j$  be the  $j$ -th Lagrange interpolation polynomial. Observe that we have

$$l_j = \prod_{i=1, i \neq j}^b (x - x_i) = \sum_{k=0}^{b-1} (-1)^k \sigma_j^k x^{b-k-1}. \quad (29)$$

**Proposition 19.** *The bias  $\gamma$  of  $p$  w.r.t  $\bar{p}$  is bounded by*

$$\max \left( \frac{\sum_{i=1}^b |\sigma_i^k x_i|}{|\sigma^{k+1}|}, k = 0, \dots, b-1 \right) \rho. \quad (30)$$

We define

$$\varpi_k = \frac{\sqrt{3 \sum_{i=1}^b |\sigma_i^k x_i|^2}}{3|\sigma^{k+1}|} \rho \quad (31)$$

$$\omega = \max(\varpi_k, k = 0, \dots, b-1). \quad (32)$$

If  $b$  is big enough, then  $\gamma$  is bounded by the normal distribution  $N(0, \omega)$ . (For the precise meaning of the statement bounded by a distribution, please refer to [Definition 27](#) in the [Appendix](#).)

**Proof.** By the definitions of  $\bar{p}$  and  $p$ , we have

$$\begin{aligned} \bar{p} - p &= \prod_{i=1}^b (x - x_i - x_i \delta_i) - \prod_{i=1}^b (x - x_i) \\ &= \prod_{i=1}^b (x - x_i) - \sum_{i=1}^b \prod_{j=1, j \neq i}^b (x - x_j) x_i \delta_i + O(\delta^2) - \prod_{i=1}^b (x - x_i) \\ &= - \sum_{i=1}^b l_i x_i \delta_i + O(\delta^2) \\ &\approx - \sum_{i=1}^b \left( \sum_{k=0}^{b-1} (-1)^k \sigma_i^k x_i \delta_i \right) x^{b-k-1} \\ &= - \sum_{k=0}^{b-1} (-1)^k \left( \sum_{i=1}^b \sigma_i^k x_i \delta_i \right) x^{b-k-1}, \end{aligned}$$

and

$$p = \prod_{i=1}^b (x - x_i) = \sum_{k=-1}^{b-1} (-1)^{k+1} \sigma^{k+1} x^{b-k-1}.$$

Thus, the absolute value of the bias for each coefficient  $\gamma_k$ , for  $k = 0, \dots, b - 1$ , is given by

$$|\gamma_k| = \frac{|\sum_{i=1}^b \sigma_i^k x_i \delta_i|}{|\sigma^{k+1}|} \leq \frac{\sum_{i=1}^b |\sigma_i^k x_i|}{|\sigma^{k+1}|} \rho.$$

Hence, to order  $O(\delta^2)$

$$\gamma \leq \max \left( \frac{\sum_{i=1}^b |\sigma_i^k x_i|}{|\sigma^{k+1}|}, k = 0, \dots, b - 1 \right) \rho.$$

Recall that, by assumption, the random variables  $\delta_1, \dots, \delta_b$  are independent. Also observe that, to order  $O(\delta^2)$ , the bias of each coefficient of  $p$  is a linear combination of these variables. Hence, we can compute the variance  $\omega_k^2$  of the bias  $\gamma_k$  of the coefficient  $x^{b-k-1}$ , for  $k = 0, \dots, b - 1$ , by means of the properties given in the [Appendix](#):

$$\begin{aligned} \omega_k^2 &= \text{Var} \left( \sum_{i=1}^b \sigma_i^k x_i \delta_i / \sigma^{k+1} \right) \\ &= \text{Var} \left( \sum_{i=1}^b \sigma_i^k x_i \delta_i \right) / |\sigma^{k+1}|^2 \\ &= \frac{\sum_{i=1}^b |\sigma_i^k x_i|^2}{|\sigma^{k+1}|^2} \text{Var}(\delta_i) \\ &\leq \frac{\sum_{i=1}^b |\sigma_i^k x_i|^2}{3|\sigma^{k+1}|^2} \rho^2 \\ &= \varpi_k^2. \end{aligned}$$

When  $b$  is big enough, the distribution of  $\gamma_k$  will tend to a normal distribution  $N(0, \omega_k)$ , by the results in the [Appendix](#). Let  $\omega = \max(\varpi_k, k = 0, \dots, b - 1)$ ; then  $\gamma_k$  is bounded by  $N(0, \omega)$  for each  $k$ . Finally,  $\gamma$  is bounded by  $N(0, \omega)$ .  $\square$

**Remark 20.** If  $\gamma$  follows the normal distribution  $N(0, \omega)$  and  $x = 2\omega$  then we have  $P(|\gamma| < x) \approx 0.95$ . In fact, our experiments show that for  $b \geq 10$ , the probability  $P(|\gamma| < x)$  is close to 0.95. Thus we can use Formula (31) to estimate the bias in the coefficients even if  $b$  is not very big. From the output of PHC we can estimate  $\delta$  using condition numbers, compute  $\omega$ , and finally estimate the bias for the coefficients with confidence level 0.95. In this section assuming  $b$  is big enough, then we have:

**Proposition 21.** Given  $n$  univariate polynomials,  $p_i(x_i) = \sum_k a_{i,k} x_i^k$ ,  $i = 1, \dots, n$ , if each  $\delta_i$  (the bias of  $p_i$ ) satisfies  $N(0, \omega)$ , then the bias of  $\prod_{i=1}^n p_i$  is bounded by  $N(0, \sqrt{n}\omega)$  to order  $O(\delta^2)$ .

**Proof.** Write the product of the univariate polynomials as a sum of monomials:

$$p_1 \cdots p_n = \sum f_e X^e,$$

where

$$f_e = f_{e_1, \dots, e_n} = a_{1,e_1} \cdots a_{n,e_n}.$$

Denote the exact coefficient by

$$\bar{f}_e = (a_{1,e_1} + a_{1,e_1}\delta_1) \cdots (a_{n,e_n} + a_{n,e_n}\delta_n).$$

By the same arguments as above,

$$\begin{aligned} \gamma_e &= \frac{\bar{f}_e - f_e}{f_e} \\ &= \frac{a_{1,e_1} \cdots a_{n,e_n} (\delta_1 + \cdots + \delta_n)}{a_{1,e_1} \cdots a_{n,e_n}} + O(\delta^2) \\ &\approx \delta_1 + \cdots + \delta_n. \end{aligned}$$

Because each  $\delta_i$  satisfies  $N(0, \omega)$ , their sum is also normally distributed (see the Appendix) with distribution function  $N(0, \sqrt{n}\omega)$ . So, to order  $O(\delta^2)$  the bias of  $\prod_{i=1}^n p_i$  is bounded by  $N(0, \sqrt{n}\omega)$ .  $\square$

**Proposition 22.** Let  $p_i(X) = \sum f_{i,e} X^e$ ,  $i = 1, \dots, N$ , be multivariate polynomials such that  $\delta_i$  (the bias of  $p_i$ ) is normally distributed with distribution  $N(0, \omega)$ . Let

$$\begin{aligned} \omega_e &= \frac{\sqrt{\sum_{i=1}^N f_{i,e}^2}}{|\sum_{i=1}^N f_{i,e}|} \omega \\ \omega' &= \max(\omega_e). \end{aligned} \tag{33}$$

Then, to order  $O(\delta^2)$ , the random variable  $\gamma$  for  $\sum_{i=1}^N p_i(X)$  is bounded by  $N(0, \omega')$ .

**Proof.** Examine the coefficients of the monomials:

$$\begin{aligned} p_1 + \cdots + p_N &= \sum f_e X^e \\ f_e &= f_{1,e} + \cdots + f_{N,e}. \end{aligned}$$

Let the exact coefficient be denoted by

$$\bar{f}_e = (f_{1,e} + f_{1,e}\delta_1) + \cdots + (f_{N,e} + f_{N,e}\delta_N).$$

Again, by the same arguments, the bias  $\gamma_e$  is

$$\frac{\bar{f}_e - f_e}{f_e} = \frac{f_{1,e}\delta_1 + \cdots + f_{N,e}\delta_N}{f_{1,e} + \cdots + f_{N,e}} + O(\delta^2).$$

Because each  $\delta_i$  is normally distributed by  $N(0, \omega)$ , the distribution of  $\gamma_e$  is still normal and equal to  $N(0, \omega_e)$  (see the Appendix). So  $\gamma$  for the sum is bounded by  $N(0, \omega')$  (again, see the Appendix for the meaning of bounded here).

**Definition 23.** Given an approximate triangular set  $T$  and the bias bound  $\rho$  of the approximate roots, let the *bias* of  $T$  be bounded by  $N(0, \omega)$ . Denote the *standard deviation* of  $T$  by  $sd$  where  $sd = \omega/\rho$ .

**Remark 24.** Let  $V \simeq_{\epsilon} \tilde{V}$ . Assume that  $\tilde{V}$  satisfies the strong equivalence condition with tolerance  $\epsilon$ , in the sense of Definition 13. Then, it follows from Propositions 19, 21 and 22 that we can determine  $sd$  and the bias of the approximate triangular sets (in the approximate equiprojective decomposition) of  $\tilde{V}$  with a given probability. Moreover, for an approximate system, given a perturbation of the approximate roots, we can estimate the change of the coefficients of the associated approximate triangular sets.

For further computations, using the approximate triangular sets will likely be difficult because of accumulation of errors. However our discussion above also provides a statistical way to estimate this accumulation.

## 5. An illustrative example

Here, we use a simple example to illustrate our concept of an approximate triangular set and our algorithm for determining the standard deviation. Let us consider

$$\text{sys} = [zx^2 - zy, x^2 - 4y + y^2 + 2, -3zy + zy^2 + 3z - 3]. \quad (34)$$

The exact triangular set of this system with order  $z < y < x$  is

$$[z - 3, y^2 - 3y + 2, x^2 - y]. \quad (35)$$

- (1) Solving the system using PHC, we get four isolated points:

$$\begin{aligned} &[z = 3.0, y = 2.0, x = 1.41421356237309, rco = 0.01511], \\ &[z = 3.0, y = 1.0, x = 1.0, rco = 0.02089], \\ &[z = 3.0, y = 2.0, x = -1.41421356237309, rco = 0.01511], \\ &[z = 3.0, y = 1.0, x = -1.0, rco = 0.02089]. \end{aligned}$$

Here  $rco$  is the inverse of the condition number of the Jacobian matrix at this point.

- (2) We remark, as we did in the Introduction, that each solved form  $[z = 3.0, y = 2.0, x = 1.41421356237309]$ ,  $[z = 3.0, y = 1.0, x = 1.0]$ ,  $[z = 3.0, y = 2.0, x = -1.41421356237309]$ ,  $[z = 3.0, y = 1.0, x = -1.0]$  is an approximate triangular set.
- (3) We use the condition numbers to estimate  $d_{\max}$ :  $\delta = 1/rco \times 10^{-16} = 6.62 \times 10^{-15}$ , and call this the estimated value of  $\rho$ . For this example, we know the exact solutions, and the exact distance between roots. In particular  $\rho$  should be  $\sqrt{2} - 1.41421356237309 = 5.1 \times 10^{-15}$ . In practice we do not know the exact solution of the input system, and we can only give an estimated value for  $\rho$ . But we need to point out that this estimation works well for many examples. Comparisons are given in the next section.
- (4) By the definition of an approximate equiprojectable decomposition, the projections of the first and third points above are numerically equal since  $|2.0 - 2.0| < (2.0/0.01511 + 2.0/0.01511) \times 10^{-16}$ .

Also the projections of the first and second points are not numerically equal since  $|2.0 - 1.0| > (2.0/0.01511 + 1.0/0.02089) \times 10^{-16} = 1.8 \times 10^{-14}$ .

In the same way, we get two different projected points  $p_1 = (3.0, 2.0)$ ,  $p_2 = (3.0, 2.0)$  on the  $zy$ -plane, and there are two points on each fiber. The projections of  $p_1$ ,  $p_2$  onto the  $z$  axis are just one point  $z = 3.0$ . So the variety of  $\text{sys}$  is approximately equiprojectable. From the cardinality of the fibers, we know that the degree sequence is  $[1, 2, 2]$  with respect to the main variables of each polynomial in the triangular set. The degree sequence can be equivalently written as  $1 \cdot 2^2$ .

- (5) By formula 7, we get the approximate triangular set of  $\text{sys}$ :

$$[-0.99999999999986y + 1.0x^2, y^2 - 3.0y + 2.0, z - 3.0]. \quad (36)$$

The biggest relative error of coefficients is  $1.4 \times 10^{-14}$ . By formula (31) and (33) the standard deviation ( $sd$ ) is 2.89.

So  $sd \times \rho = 1.9 \times 10^{-14} > 1.4 \times 10^{-14}$  is a good estimate for the relative error. In the next section we will give more non-trivial examples to support our statement. Due to both input and round-off errors in numerical computation, there will be some monomials

Table 1

Experiments for our probabilistic analysis ( $sd$  = standard deviation, rel. err. is relative error)

# roots	# tests	% of trials: rel. err. > 1 $sd$ (0.32 expected)	% of trials: rel. err. > 2 $sd$ (0.05 expected)	% of trials: rel. err. > 3 $sd$ (0.003 expected)
10	1000	0.328	0.0503	0.0168
20	1000	0.312	0.0425	0.0050
30	1000	0.350	0.0579	0.0023
40	800	0.335	0.0517	0.0067
50	500	0.342	0.0474	0.0042

Table 2

Input systems ( $n$  = # of polynomials;  $d$  = degree system;  $h$  = height input coefficients;  $H$  = height output coefficients;  $\hat{H}$  = estimated height output coefficients)

Sys	Name	$n$	$d$	$h$	$H$	$\hat{H}$	Reference
1	Issac97	4	2	2	71	1498	(The Symbolicdata Project, 2000–2002)
2	L3	3	3	1	1	1678	(Aubry and Moreno Maza, 1999)
3	Sendra	2	7	7	59	2421	(The Symbolicdata Project, 2000–2002)
4	fabfaux	3	3	13	72	2650	(European Commission, 1996)
5	L4	3	4	1	2	3977	(Aubry and Moreno Maza, 1999)
6	Cylohexne	3	4	3	9	4361	(The Symbolicdata Project, 2000–2002)
7	Weispfenning94	3	5	0	10	7392	(The Symbolicdata Project, 2000–2002)
8	UteshevBikker	4	3	3	88	7908	(The Symbolicdata Project, 2000–2002)
9	Fee-1	4	2	2	34	23 967	(The Symbolicdata Project, 2000–2002)
10	Reimer-4	4	5	1	14	56 013	(The Symbolicdata Project, 2000–2002)
11	S9 <sub>1</sub>	8	2	2	33	58 116	(The Symbolicdata Project, 2000–2002)
12	eco6	6	3	0	12	105 718	(The Symbolicdata Project, 2000–2002)
13	Geneig	6	3	2	82	114 466	(The Symbolicdata Project, 2000–2002)
14	gametwo5	5	4	8	674	158 075	(The Symbolicdata Project, 2000–2002)
15	dessin-2	10	2	7	436	360 596	(The Symbolicdata Project, 2000–2002)
16	eco7	7	3	0	26	387 754	(The Symbolicdata Project, 2000–2002)
17	Methan61	10	2	16	227	452 756	(The Symbolicdata Project, 2000–2002)

of approximate triangular sets with very small coefficients that do not appear in the exact triangular sets. Then the biggest relative error of coefficients is 1. So in practice we will consider coefficients which are smaller than a given tolerance as 0.

## 6. Experimental results

We have conducted two sets of experiments. The first one illustrates the probabilistic analysis of Proposition 19. Experiments are described in Section 6.1, and the results appear in Table 1.

The second set of experiments deals with the computation of exact and approximate triangular decompositions. Section 6.2 presents the exact case whereas Section 6.3 reports on the approximate one. Most of the test polynomial systems that we use (see Table 2) are well known problems (Aubry and Moreno Maza, 1999; Dahan et al., 2005; The Symbolicdata Project, 2000–2002). They are zero-dimensional square systems defined by multivariate polynomials over  $\mathbb{Q}$  generating radical ideals. Table 3 shows data for the exact triangular decompositions of these systems, the output from PHC is collected in Table 4, and Table 5 shows the approximate

Table 3

Exact equiprojective triangular decomposition with the RegularChains library

Sys.	Exact equiproj. dec. tim. (s)	Degree configuration	# $\mathbb{C}$ -roots	Time to isolate $\mathbb{R}$ -roots (s)	# $\mathbb{R}$ -roots
1	164	16 1 <sup>3</sup>	16	<1	0
2	<1	(1 3 1), (8 1 1), (8 2 1)	27	<1	5
3	33	46 1	46	5	6
4	28	27 1 <sup>2</sup>	27	1	3
5	1	(24 2 1), (16 1 1)	64	<1	8
6	6	(4 1 2), (8 1 1)	16	<1	12
7	72	54 1 <sup>2</sup> 1	54	<1	0
8	29 201	36 1 <sup>3</sup>	36	7	10
9	24	26 1 <sup>3</sup>	26	2	6
10	10 097	18 2 1 <sup>2</sup>	36	5	4
11	26	10 1 <sup>7</sup>	10	1	4
12	50	16 1 <sup>5</sup>	16	<1	4
13	18	10 1 <sup>3</sup>	10	2	10
14	24 320	44 1 <sup>4</sup>	44	45	12
15	527	1 42 1 <sup>8</sup>	42	15	1
16	2742	32 1 <sup>6</sup>	32	4	8
17	6251	27 1 <sup>9</sup>	27	28	13

Table 4

Approximate roots from PHC where the estimate  $\rho = \text{condition number} \times 10^{-16}$  and exact  $\rho = \text{largest 2-norm of the distance between the exact and approximate roots divided by the 2-norm of the approximate root}$ 

Sys.	# $\mathbb{C}$ -roots	# $\mathbb{C}$ -roots by PHC	PHC tim. (s)	Estimated $\rho$	Exact $\rho$
1	16	16	1	0.448e-14	0.239e-14
2	27	27	1	0.186e-14	0.337e-14
3	46	46	4	0.159e-11	0.274e-14
4	27	27	2	0.224e-14	0.154e-14
5	64	64	1	0.143e-14	0.331e-14
6	16	16	<1	0.835e-14	0.181e-14
7	54	49	5	0.183e-13	0.336e-14
8	36	36	6	0.767e-12	0.781e-14
9	26	26	5	0.229e-11	0.759e-14
10	36	36	3	0.739e-13	0.544e-14
11	10	10	3	0.107e-13	0.125e-14
12	16	16	3	0.292e-13	0.287e-14
13	10	10	2	0.629e-13	0.105e-13
14	44	43	6	0.665e-12	0.144e-13
15	42	41	11	0.585e-7	0.271e-14
16	32	32	14	0.760e-13	0.264e-14
17	27	13	10	0.846e-6	0.563e-13

triangular decompositions computed from the PHC output. The main results for the purposes of this paper are given by this latter table.

### 6.1. Normal distribution test

Let  $b$  be a number of roots given in the column # roots. We randomly generate  $b$  roots, and view them as the exact roots of a polynomial  $\bar{p}$  of degree  $d$ . Then, we perturb each of these

Table 5

Approximate triangular sets:  $sd$  = standard deviation defined in Section 4;  $exact \rho$  = largest 2-norm distance between the exact and approximate roots divided by the 2-norm of the approximate root;  $\delta_{coeff}$  = largest relative error of the coefficients of the approximate triangular set compared with the exact one

Sys.	$sd$	Exact $\rho \cdot sd$	$\delta_{coeff}$	< $sd$ ?	< $2sd$ ?	Residual
1	403.3	0.9639e−12	0.197e−12	yes	yes	0.444e−15
2	7.492	0.2529e−13	0.211e−13	yes	yes	0.125e−13
3	1729.2	0.4736e−11	0.542e−11	no	yes	0.89e−11
4	1056.7	0.1625e−11	0.463e−12	yes	yes	0.201
5	59 188.4	0.1959e−09	0.248e−09	no	yes	0.555e−7
6	23 835.5	0.4314e−10	0.179e−11	yes	yes	0.7e−13
7	NA	NA	NA	NA	NA	NA
8	383.8	0.2996e−11	0.942e−12	yes	yes	0.163e−8
9	151.6	0.1151e−11	0.181e−12	yes	yes	0.504e−13
10	3928.4	0.2137e−10	0.397e−12	yes	yes	0.193e−18
11	45.77	0.5708e−13	0.133e−13	yes	yes	0.188e−15
12	121.7	0.3488e−12	0.184e−12	yes	yes	0.216
13	551.7	0.5815e−11	0.761e−13	yes	yes	0.314e−17
14	NA	NA	NA	NA	NA	NA
15	NA	NA	NA	NA	NA	NA
16	317.7	0.8397e−12	0.154e−11	no	yes	0.218e20
17	NA	NA	NA	NA	NA	NA

roots by a uniformly distributed random variable, leading to an approximate polynomial  $p$ . The two polynomials  $\bar{p}$  and  $p$  are expanded in order to obtain  $\varepsilon$ , the largest relative error for a coefficient. We compute the standard deviation  $sd$  using formula (31), and compare it with  $\varepsilon$ . These experiments are repeated many times (between 500 and 1000; see the column # tests) for  $b = 10, 20, 30, 40, 50$ . The third column is the percentage of times for which the relative error is bigger than one standard deviation. If the relative error is normally distributed, then this percentage should be 0.32, which we verify in our tests.

## 6.2. Exact triangular decomposition

The test polynomial systems are given in Table 2. For each input system  $F$ , we give  $n$  the number of variables,  $d$  the total degree of  $F$ , the logarithm  $h$  of the largest coefficient, the number of digits  $H$  appearing in the largest coefficient in the (exact) equiprojective decomposition of  $F$ , and the height  $\widehat{H}$  of that coefficient as estimated using the formulas of Dahan et al. (2005).

In order to compute the exact equiprojective decomposition, we use the RegularChains library written in MAPLE by Lemaire, Moreno Maza and Xie (Lemaire et al., 2005) in which the algorithms of Moreno Maza (2000), Dahan et al. (2005) are implemented. Our computations are done on a 2799 MHz Pentium 4 machine. The timings for computing the exact equiprojective decompositions are given in the first column of Table 3. To understand these timings, we should mention that the RegularChains code is high-level interpreted code (and not compiled). Moreover, this code is not supported by fast arithmetic, such as FFT-based arithmetic.

Each degree configuration specifies the degree sequences of the triangular sets in the decomposition (see Aubry and Moreno Maza (1999) for similar data). Hence, the number of sequences in a degree configuration equals the number of equiprojective components of the system. In Table 3, #  $\mathbb{C}$ -roots and #  $\mathbb{R}$ -roots are, respectively, the total numbers of complex and real roots of the system. The column labeled “Time to isolate  $\mathbb{R}$ -roots”, gives the total time in seconds taken to isolate all the real roots to a precision of  $2^{-30}$  using interval arithmetic.

We have also isolated each complex root. This was done by Éric Schost (École Polytechnique, France) using *Magma* as follows. First, the *splitting circle* method of Schönage was used to separate the complex roots. Then, Newton iteration was used to refine the isolation boxes. A precision of 200 digits could be achieved for our 17 test systems in less than 10 min on a Pentium P3 running at 1 GHz.

### 6.3. Approximate triangular sets

We used the PHC package (Verschelde, 1999; Leykin and Verschelde, 2004) to compute the approximate isolated roots for each benchmark system. Then we interpolated the approximate triangular sets and we give the results of our error analysis for each system. The computations in Tables 4 and 5 were done on a 1.5 GHz Pentium M machine, and the timings for finding the roots using PHC are listed in *PHC Timing* of Table 4. In Table 4: The first column is the exact number of roots and second column is the number of roots found by PHC. For some systems, PHC (in black box mode) did not get every root. This simply means that the default settings in the black box version of PHC did not solve the system. We did not compute the approximate triangular sets for such systems. Some of these systems could certainly have been solved by using PHCPack, by exploiting the flexibility of its powerful user specified options, designed for more challenging problems. But we did not do that here. The *estimate*  $\rho$  is defined as the condition number  $\times 10^{-16}$ , and *exact*  $\rho = \max(|x_i - \bar{x}_i|/|x_i|)$ ,  $\bar{x}_i \in V$ , where the  $\bar{x}_i$  are the “exact” roots, the  $x_i$  are the roots given by PHC, and the distance is given by the 2-norm. The results show that our estimated distance is often larger than the exact distance.

In Table 5: The second column gives the standard deviation of the approximate triangular set, as discussed in Remark 24. The third column is the product of the exact  $\rho$  and one standard deviation. In the fourth column  $\delta_{\text{coeff}}$  is the largest relative error of the coefficients of the approximate triangular set as compared with the exact one. If this relative error is less than exact  $\rho \cdot sd$ , the element of the fifth column (labeled  $<sd?$ ) is “yes”; otherwise it is “no”. Moreover, for every approximate triangular set, the relative error is bounded by  $2sd$  (see column 6). The last column, labeled *Residual*, gives the maximum residual of an approximate triangular set at the roots given by PHC. The results of this table support the conclusions of Remark 24.

## 7. Discussion

There are well-developed algorithms for computing exact triangular decompositions and considerable recent improvements in their time complexity (Dahan et al., 2005). Such representations are desirable, not only because of their triangular solved-form structure, but also because, in comparison to other exact methods, their space complexity is well controlled (Dahan and Schost, 2004). In particular, they use the minimum number of polynomials needed to describe the equi-dimensional components of a polynomial system.

We have extended such methods to approximate systems in the dimension zero case. We have exploited methods from the newly developing area of Numerical Algebraic Geometry (Sommese and Verschelde, 2000; Verschelde, 1999; Sommese et al., 2001; Sommese and Wampler, 2005), together with new techniques based on the so-called equiprojective decomposition (Dahan et al., 2004) of a zero-dimensional variety.

Throughout this paper we have assumed that the input is zero dimensional and generates a radical ideal. We briefly discuss the situation where both of these restrictions are removed. The approximate methods in Sommese and Wampler (2005), Verschelde (1999) yield isolated

points, possibly of higher multiplicity, corresponding to the zero-dimensional equi-dimensional components. Such multiplicities can be removed (deflated) numerically using the techniques of Dayton and Zeng (2005) and Leykin et al. (2005) (see Lecerf (2002) for a symbolic method for the exact case) and subsequently the methods of our paper can be applied.

Our contribution, in the zero-dimensional case, has been to show that the isolated points, given by approximate coordinates, can be interpolated in order to obtain a triangular decomposition which is an approximation of the exact equiprojective decomposition. The methods (Sommese et al., 2001) yield a numerical irreducible decomposition for this case, and in particular they give a collection of triangular sets, each of them corresponding trivially to an isolated point.

In addition, the co-dimension one components (hypersurfaces) can be numerically interpolated by Sommese and Verschelde (2000), Sommese et al. (2001) to obtain a single polynomial which can be considered as a representation with triangular shape. The methods also give (non-triangular) representations of all of the positive dimensional components using generic points on each component. The above results, together with those in our paper on linearized triangular decompositions (Moreno Maza et al., 2007), represent progress on the general problem of obtaining approximate triangular representations for all components of a given polynomial system.

Often, in applications, polynomial systems have parameters (Sommese and Wampler, 2005). One is interested in behavior at generic values of the parameters. In practice, one proceeds by selecting generic values for the parameters, and this is often how zero-dimensional polynomial solving arises in applications. In Sommese and Wampler (2005), it is shown how once a solution is computed by homotopy continuation for a specific parameter value, then solutions for other parameter values can be obtained efficiently from the given one using a “parameter homotopy”. Analogously, we can follow this idea to reduce positive dimensional systems to zero-dimensional ones by setting generic values for the parameters. Then, a parameter homotopy is used to efficiently compute approximate triangular sets for other parameter values. A promising approach for constructing triangular sets of positive dimensional components is to use parameter homotopies followed by interpolation by choosing sufficiently many values for the parameters. Thus, our work on the zero-dimensional case is a preparation for the study of the general case. The related exact approaches go back to Van der Waerden (1953), Gianni et al. (1988) among others; also see the recent work Dahan et al. (2006), Schost (2003).

Under some choices of interpolation points (e.g. uniformly spread points) the interpolation formulas of Dahan and Schost (2004) may be ill-conditioned (Berrut and Trefethen, 2004; Higham, 2004). In the zero-dimensional case, we have no control over this, since the locations of the points are fixed. However, the stability analysis of our paper can identify this situation. In particular, a very large standard deviation means that the coefficients are very sensitive to changes in the roots. For such systems, interpolation is not a good method for obtaining approximate triangular sets from the roots.

In Chèze and Galligo (2005), the authors compute an exact absolute factorization of a bivariate polynomial from an approximate factorization. It is natural to ask whether one could compute an exact equiprojective decomposition from an approximate one. One preliminary answer is as follows. Let  $F$  be an (exact) zero-dimensional polynomial system in  $\mathbb{Q}[X_1 \prec \dots \prec X_n]$  with total degree  $d$  and the maximum number of digits of the coefficients  $h$ . Then (Dahan et al., 2005), the height of any coefficient of any (exact) triangular set in the equiprojective decomposition of  $V(F) \subseteq A^n(\mathbb{C})$  is  $O(h n d^n)$ . This suggests that the numbers  $d$  and  $n$  must be small for this *reconstruction* (from approximate to exact) to be realistic. However, the question remains open for future work. Indeed, Table 2 shows that the actual coefficient size  $H$  in the triangular

set is much less than the above height upper bound  $\widehat{H}$ . Another approach is to lift to nearby exact triangular systems which may have moderately sized rational coefficients, in comparison to lifting to exact rational triangular systems. In addition, a linearized sensitivity analysis should yield information on coefficient versus solution changes (e.g. see Stetter (2004)). This information is valuable in lifting exact results from the approximate triangular decomposition. Such approaches are a topic of future work.

Traditional uses of exact triangular sets include finding the reduced or simplified form of a polynomial with respect to a triangular decomposition, as accomplished by a chain of pseudo-reductions. Standard deviations of the coefficients also provide information about the accumulation of error in such operations. Provided that the chains of reductions are short, and the degrees of the polynomials involved are not too high, some similar uses are possible with our approximate triangular systems. However, we caution the reader that the accumulation of round-off error means that such operations should be carried out with care.

The roots of a generic zero-dimensional system are equiprojectable and correspond to a normalized triangular set. Following the idea in Stetter (2004), we can construct a homotopy to study the deformations of triangular sets with special shape (using the *Shape-Lemma*) and the errors in the roots caused by errors in the coefficients. This idea will also be pursued in future work.

Finally, we direct the reader to Stetter (1999, 2004), where fundamental theorems on backward error analysis and sensitivity of the roots under small perturbations of the coefficients are given for polynomials. When the input system  $F$  is approximate, although discontinuous phenomena can occur, some continuity aspects are preserved under perturbation (Stetter and Thallinger, 1998).

The favorable properties of the equiprojectable decomposition of  $V(F)$  under specialization (Dahan et al., 2005) suggest that the continuity of approximate equiprojectable decomposition needs to be studied in future application to general systems.

## Acknowledgements

GR and WW gratefully acknowledge support for this work from the NSF funded IMA Thematic Year on Applications of Algebraic Geometry. All of the authors acknowledge support from the National Sciences and Research Engineering Council of Canada.

The authors would like to thank Yuzhen Xie (University of Western Ontario, Canada) who realized the experiments with the `RegularChains` library (Lemaire et al., 2005). And we also thank the following colleagues for their assistance: François Lemaire (University of Lille 1, France) who provided the source code of `Realroots` for computing the real solutions with the `RegularChains` library and Éric Schost (École Polytechnique, France) who isolated each complex root using `Magma`.

## Appendix — Brief review of probability theory

In our stability analysis of coefficients, a probability model was introduced. Here we give a brief review of the relevant standard probability knowledge required.

- If  $\delta$  is a random variable and  $c$  is a constant in  $\mathbb{R}$  then  $\text{Var}(c\delta) = c^2\text{Var}(\delta)$ .
- If  $\delta_1, \dots, \delta_b$  are random variables and  $\xi = \sum \delta_i$  then the expectation value is additive:  $E(\xi) = \sum E(\delta_i)$ . Moreover, if they are independent, then the variance of the sum of these random variables is also additive:  $\text{Var}(\xi) = \sum \text{Var}(\delta_i)$ .

- Suppose  $\delta = \delta_{re} + \delta_{im}\sqrt{-1}$  and  $\delta_{re}, \delta_{im}$  are independent random variables with the same distribution with  $c \in \mathbb{C}$ . Then  $\text{Var}(\Re(c\delta)) = |c|^2\text{Var}(\delta_{re}) = \text{Var}(\Im(c\delta)) = |c|^2\text{Var}(\delta_{im})$ , where  $\Re(z)$  and  $\Im(z)$  are the real and imaginary parts of  $z$ . In this paper we define  $\text{Var}(\delta) := \text{Var}(\delta_{re})$ .
- $N(0, 1)$  is the standard normal distribution with mean 0, standard deviation 1, probability density function  $p(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$  and cumulative density function  $\Phi(x) = \int_{-\infty}^x p(x)dx$ . Note that  $\Phi(1) \approx 0.68$ ,  $\Phi(2) \approx 0.95$ .
- Suppose that  $\delta_1, \dots, \delta_b$  are independent random variables with distribution functions  $F_1, \dots, F_b$  and  $E(\delta_i) = 0$ ,  $0 < \text{Var}(\delta_i) < \infty$ ,  $s_b^2 = \sum \text{Var}(\delta_i)$ . The Lindeberg condition for a sum of independent random variables is that for any  $t > 0$ :

$$\frac{1}{s_b^2} \sum_{k=1}^b \int_{|x|>ts_b} x^2 dF_k(x) \longrightarrow 0 \quad \text{when } b \longrightarrow \infty. \quad (37)$$

From our assumptions about the roots, the bias is uniformly distributed and because  $0 < \text{Var}(\delta_i) < \infty$  we have  $s_b^2 \rightarrow \infty$  as  $b \rightarrow \infty$ . So for any  $t > 0$ , there always exists an  $L$ , when  $b > L$  the integral above is 0.

**Proposition 25** (*Uniform Distribution and Lindeberg Condition*). *If  $\delta_1, \dots, \delta_b$  are independent random variables which are uniformly distributed, and  $E(\delta_i) = 0$ , if the variance of each  $\delta_i$  is non-zero and finite, then this family of random variables satisfies the Lindeberg condition.*

**Proposition 26** (*Lindeberg's Central Limit Theorem (Shiryayev, 1995)*). *Suppose  $\delta_1, \dots, \delta_b$  are uniformly distributed independent random variables,  $E(\delta_i) = 0$  and  $\delta_i$  satisfies the Lindeberg condition. Let  $S_b = \sum_{i=1}^b \delta_i$  and  $s_b^2 = \sum_{i=1}^b \text{Var}(\delta_i)$  then when  $b \rightarrow \infty$ , the sum of variables divided by its standard deviation is convergent (in distribution) to a standard normal distribution:*

$$\frac{S_b}{s_b} \longrightarrow N(0, 1) \quad \text{as } b \longrightarrow \infty. \quad (38)$$

**Definition 27.** We say a random variable  $\xi$  or  $|\xi|$  is bounded by  $N(0, \omega)$  if the probability  $P(|\xi| < x\omega) > \Phi(x)$ .

When  $\omega$  is bigger, the probability will also be bigger. In particular if  $\omega' > \omega$  then  $P(|\xi| < x\omega') > P(|\xi| < x\omega)$ , so  $\xi$  is also bounded by  $N(0, \omega')$ .

## References

- Aubry, P., Lazard, D., Moreno Maza, M., 1999. On the theories of triangular sets. *J. Symbolic Comput.* 28 (1–2), 45–124.  
 Aubry, P., Moreno Maza, M., 1999. Triangular sets for solving polynomial systems: A comparative implementation of four methods. *J. Symbolic Comput.* 28 (1–2), 125–154.  
 Aubry, P., Valibouze, A., 2000. Using Galois ideals for computing relative resolvents. *J. Symbolic Comput.* 30 (6), 635–651.  
 Berrut, J.P., Trefethen, L.N., 2004. Barycentric Lagrange interpolation. *SIAM Rev.* 46 (3), 501–517.  
 Blum, L., Cucker, F., Shub, M., Smale, S., 1997. Complexity and Real Computation. Springer, New York.  
 Chèze, G., Galligo, A., 2005. From an approximate to an exact absolute polynomial factorization. Technical report. Université de Nice.  
 Chou, S.C., 1988. Mechanical Geometry Theorem Proving. D. Reidel Publ. Comp., Dordrecht.  
 Dahan, X., Jin, X., Moreno Maza, M., Schost, É., 2006. Change of ordering for regular chains in positive dimension. In: Maple Conference. Canada, pp. 26–34.  
 Dahan, X., Moreno Maza, M., Schost, É., Wu, W., Xie, Y., 2004. Equiprojective decompositions of zero-dimensional varieties. In: ICPSS. University of Paris 6, France, pp. 69–71.

- Dahan, X., Moreno Maza, M., Schost, É., Wu, W., Xie, Y., 2005. Lifting techniques for triangular decompositions. In: ISSAC'05. ACM Press, pp. 108–115.
- Dahan, X., Schost, É., 2004. Sharp estimates for triangular sets. In: ISSAC 04. ACM Press, pp. 103–110.
- Dayton, Barry H., Zeng, Zhonggang, 2005. Computing the multiplicity structure in solving polynomial systems. In: ISSAC 05. ACM Press, pp. 116–123.
- European Commission, 1996. FRISCO — A Framework for Integrated Symbolic/Numeric Computation. Esprit Scheme Project No. 21 024. <http://www.nag.co.uk/projects/FRISCO.html>.
- Gao, X.S., Luo, Y., November 2004. A characteristic set algorithm for difference polynomial systems. In: Int. Conf. Poly. Sys. Sol., Paris, pp. 24–26.
- Gianni, P., Trager, B., Zacharias, G., 1988. Gröbner bases and primary decomposition of polynomial ideals. *J. Symbolic Comput.* 6 (2–3), 149–167.
- Higham, N.J., 2004. The numerical stability of barycentric Lagrange interpolation. *IMA J. Numer. Anal.* 24 (547–546).
- Janovitz-Freireich, Itnuit, Rónyai, Lajos, Szántó, Ágnes, 2006. Approximate radical of ideals with clusters of roots. In: ISSAC 06. ACM Press, pp. 146–153.
- Kalkbrenner, M., 1993. A generalized Euclidean algorithm for computing triangular representations of algebraic varieties. *J. Symbolic Comput.* 15, 143–167.
- Lazard, D., 1992. Solving zero-dimensional algebraic systems. *J. Symbolic Comput.* 13, 117–133.
- Lecerf, G., 2002. Quadratic Newton iteration for systems with multiplicity. *Found. Comput. Math.* 2, 247–293.
- Lemaire, F., Moreno Maza, M., Xie, Y., 2005. The RegularChains library. In: Maple Conference. Canada, pp. 355–368.
- Leykin, A., Verschelde, J., Zhao, A., 2005. Evaluation of jacobian matrices for Newton's method with deflation to approximate isolated singular solutions of polynomial systems. In: SNC 2005 Proceedings, pp. 19–28.
- Leykin, A., Verschelde, J., 2004. Phcmaple: A maple interface to the numerical homotopy algorithms in phcpack. In: Proceedings of ACA'04. University of Texas at Beaumont, USA, pp. 139–147.
- Moreno Maza, M., 2000. On triangular decompositions of algebraic varieties. Technical Report 4/99. NAG, UK. In: MEGA-2000 Conference, Bath, UK. <http://www.csd.uwo.ca/~moreno>.
- Moreno Maza, M., Reid, G., Scott, R., Wu, W., 2007. On approximate linearized triangular decompositions. In: Wang, D.M., Zhi, L. (Eds.), *Symbolic–Numeric Computation*. Birkhäuser, Basel, Boston, pp. 268–287.
- Ritt, J.F., 1932. *Differential Equations from an Algebraic Standpoint*, vol. 14. American Mathematical Society, New York.
- Schost, É., 2003. Complexity results for triangular sets. *J. Symbolic Comput.* 36 (3–4), 555–594.
- Shiryayev, A.N., 1995. *Probability*. Springer.
- Sommese, A.J., Verschelde, J., 2000. Numerical homotopies to compute generic points on positive dimensional algebraic sets. *J. Complexity* 16 (3), 572–602.
- Sommese, A.J., Verschelde, J., Wampler, C.W., 2001. Numerical decomposition of the solution sets of polynomial systems into irreducible components. *SIAM J. Numer. Anal.* 38 (6), 2022–2046.
- Sommese, A.J., Wampler, C.W., 2005. *The Numerical Solutions of Systems of Polynomials*. World Scientific, Singapore.
- Stetter, Hans J., 1999. The nearest polynomial with a given zero, and similar problems. *SIGSAM Bull.* 33 (4), 2–4.
- Stetter, Hans J., 2004. *Numerical Polynomial Algebra*. SIAM, Philadelphia.
- Stetter, Hans J., Thallinger, Gunther H., 1998. Singular systems of polynomials. In: ISSAC'98. ACM Press, pp. 9–16.
- The SymbolicData Project, 2000–2002. <http://www.SymbolicData.org>.
- Verschelde, J., 1999. PHC pack: A general-purpose solver for polynomial systems by homotopy continuation. *ACM Trans. Math. Software* 25 (2), 251–276.
- Van der Waerden, B.L., 1953. *Modern Algebra*, 2nd ed. Frederick Ungar Publishing Co., New York.
- Wang, D.M., 2000. *Elimination Methods*. Springer, Wein, New York.
- Wu, W.T., 1987. A zero structure theorem for polynomial equations solving. *MM Research Preprints* 1. pp. 2–12.