Pivoting in Extended Rings for Computing Approximate Gröbner Bases

Jean-Charles Faugère and Ye Liang

Abstract. It is well known that in the computation of Gröbner bases arbitrarily small perturbations in the coefficients of polynomials may lead to a completely different staircase, even if the solutions of the polynomial system change continuously. This phenomenon is called *artificial discontinuity* in Kondratyev's Ph.D. thesis. We show how such phenomenon may be detected and even "repaired" by using a new variable to rename the leading term each time we detect a "problem". We call such strategy the TSV (Term Substitutions with Variables) strategy. For a zero-dimensional polynomial ideal, any monomial basis (containing 1) of the quotient ring can be found with the TSV strategy. Hence we can use TSV strategy to relax term order while keeping the framework of Gröbner basis method so that we can use *existing* efficient algorithms (for instance the F_5 algorithm) to compute an approximate Gröbner basis. Our main algorithms, named TSVn and TSVh, can be used to repair artificial ε-discontinuities. Experiments show that these algorithms are effective for some nontrivial problems.

Mathematics Subject Classification (2000). Primary I.1.2; Secondary F.2.1.

Keywords. Approximate Gröbner basis, artificial discontinuity, monomial basis, F_5 algorithm.

1. Introduction

Consider a toy system

$$f_1 := 4x^2 + y^2 - 4,$$
$$f_2 := 4\varepsilon xy + 15y^2 - 12,$$

where ε is a "small" number (possibly zero). We want to compute the Gröbner basis of $\langle f_1, f_2\rangle$ w.r.t. the degree reverse lexicographic order (DRL) $\preceq$ with $y \prec x$. Buchberger's algorithm first computes the S-polynomial of f_1 and f_2. For this, we need to determine $\mathrm{lm}(f_2)$, the leading monomial of f_2 w.r.t. $\preceq$. Note that if $\varepsilon \neq 0$, then $\mathrm{lm}(f_2) = 4\varepsilon xy$ else $\mathrm{lm}(f_2) = 15y^2$. Thus, the computation needs to be branched. We obtain a Gröbner basis $G_{\varepsilon\neq 0} = \{xy + 15y^2/(4\varepsilon) - 3/\varepsilon, x^2 + y^2/4 - 1, y^3 + 48\varepsilon x/(225 + 4\varepsilon^2) -$

$(180+16\varepsilon^2)y/(225+4\varepsilon^2)\}$ or $G_{\varepsilon=0}=\{x^2-4/5, y^2-4/5\}$. Look at the following picture:

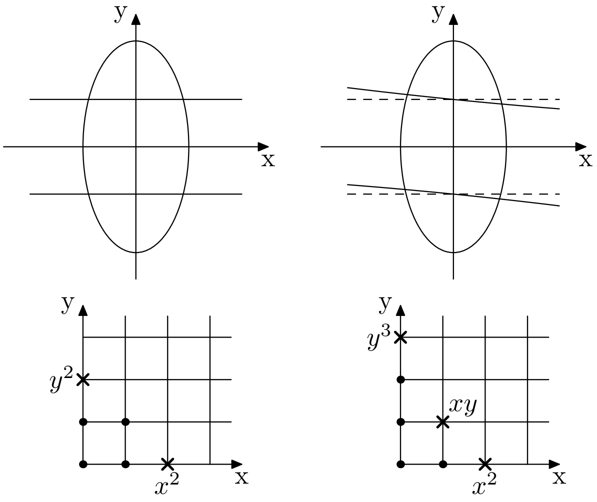

FIGURE 1. Solutions and staircases of $\langle f_1, f_2\rangle$ for $\varepsilon=0$ and $\varepsilon=0.75$

The solutions of $f_1=f_2=0$ are continuous on ε; however, we see that the staircases are different. This phenomenon is called "representation singularity" in Stetter's paper [35], and "artificial discontinuity" in Kondratyev's Ph.D. thesis [20] supervised by Stetter and Winkler. We follow the latter.

We propose a way to "repair" artificial discontinuities and explain it by using the above toy example. The key idea is simple: Rename xy with a new variable z, keep the DRL order and make $z \prec y \prec x$. Then we have a new polynomial $f_3 := 15y^2+4\varepsilon z-12$ with $\mathrm{lm}(f_3)=15y^2$. No branching is needed anymore. Yet we need to add the definition of the new variable by $f_4 := z-xy$. Continue the computation and we obtain the final Gröbner basis G as

$$\left\{\begin{array}{c} xy-z, \\ 15y^2+4\varepsilon z-12, \\ 4x^2+y^2-4, \\ (225+4\varepsilon^2)yz-180x+48\varepsilon y, \\ (225+4\varepsilon^2)z^2+36\varepsilon z-144, \\ 15xz-\varepsilon yz-12y \end{array}\right\}.$$

The fact that $\mathrm{lc}(g)|_{\varepsilon=0}\neq 0$ for every $g\in G$ implies that G is a uniform minimal Gröbner basis w.r.t. $\preceq$ for all small ε. Then we say that the artificial discontinuity is *locally repaired* by the TSV (Term Substitutions with Variables) strategy. Note that the extended ideal and the original ideal are equivalent in the sense of three aspects: monomial bases of their quotient rings (Theorem 3.4), ideal membership (Proposition 4.1) and essential zeros (Proposition 4.2). Hence, we can use the Gröbner basis of the extended ideal to solve problems on these aspects of the original ideal safely. In other words, though we do not always compute a Gröbner basis of the original ideal, we can still find monomial bases of the original quotient ring, determine the membership of the original ideal and work out the solutions of the original polynomial system by using the Gröbner basis of an extended ideal w.r.t. some term order in an extended ideal. Moreover, we set up a theory of "artificial discontinuity" of single-parametric Gröbner bases in [16] based on the techniques in this paper (see also [15]).

The "approximate Gröbner basis" in this paper means a certain generalization of the concept of Gröbner basis in numerical computation, i.e., at least one algorithm to compute it can be obtained by modifying an algorithm for Gröbner bases. Obviously, in numerical computation, the result obtained above is a kind of approximate Gröbner basis, though the ring has changed.

There is a vast literature on approximate Gröbner bases. We briefly review some of the most significant contributions in this area:

Stetter and Kondratyev used a new variable e with $e \prec 1$ to multiply $t = \mathrm{lt}(f)$ [35, 20] such that $e \cdot t \prec 1$, where $f \in \mathbb{K}[x_1, \ldots, x_n]$ with small $\mathrm{lc}(f)$. Then the leading term of f is changed. The changed order does not keep a term order, and the result is generally not a Gröbner basis.

Border basis [4, 21, 23, 36, 19, 17, 18, 9, 3] gains more freedom for finding a basis of $\mathbb{K}[x_1, \ldots, x_n]/I$ than Gröbner basis [19]. There is still a constraint that the monomial basis obtained must be closed (cf. Definition 2.7).

Mourrain and Trébuchet relaxed term orders and computed "generalized normal forms" [24, 38]. The monomial bases obtained by their method need not be a closed set but must be "connected to 1" [24].

Greg Reid, Lihong Zhi *et al.* computed a polynomial basis of $\mathbb{K}[x_1, \ldots, x_n]/I$ [26, 25]. This method is based on geometric theory of partial differential equations and is not Gröbner-like.

Besides, Shirayanagi, Sasaki and some other authors focused on numerical techniques [33, 34, 37, 39, 29, 30, 31, 32]. They did not change the theory of Gröbner basis, but traced the Gröbner basis computation and tried to avoid large loss of accuracy of the coefficients of polynomials.

The main goal of this paper is to use the TSV strategy to relax term order while keeping the framework of the Gröbner basis method. One benefit of this approach is that we can use *existing* efficient algorithm (for instance the F_5 algorithm [13]) to compute an approximate Gröbner basis. In this paper, we show that, from a theoretical point of view, given an arbitrary degree term order, *any monomial* basis (containing 1) of the quotient ring $\mathbb{K}[x_1, \ldots, x_n]/I$ can be obtained by the TSV strategy (cf. Theorem 3.4). In other words, it means that, in some sense, we have the theoretical power to simulate dedicated algorithms for computing approximate Gröbner basis (for instance border basis method or generalized normal forms) using slightly modified versions of existing standard algorithms.

Moreover, we can get a great freedom to avoid artificial ε-discontinuities (cf. Section 5): we can balance the coefficients of polynomials arising during the computation by *variable scaling strategy*, i.e., using a number d to scale some variable each time when we detect a problem. So that some coefficients of the unitized polynomials would not have small absolute values (see Section 5 for the exact definitions of “small” and “unitization”).

We provide algorithms ε-Buchbeger and ε-MatrixF5 to compute Gröbner bases and test ε-discontinuities as well as algorithms TSVn and TSVh to implement the TSV strategy. We have done first implementations of these algorithms in Maple. Although not as efficient as the procedures with C code of FGb Maple package [1, 12] or in Magma, they are sufficient to analyze the properties of the computed Gröbner bases. It will be an object of future work to compare the practical behavior with other methods.

In the following parts, Section 2 contains necessary notions and notations; Sections 3 and 4 are devoted to presenting theoretic results; Section 5 provides algorithms to test and deal with artificial ε-discontinuities; Section 6 shows experimental results; and finally we make a conclusion and state the tendency of the future work in Section 7.

2. Extended Ideal and the TSV Strategy

This section is about necessary notions and notations used in this paper.

Let $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_s\}$ be two finite sets of indeterminates with $X \cap Y = \emptyset$, $\mathbb{K}$ be field $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$. Let I be an ideal of $\mathbb{K}[X]$, T^X be the set of terms in X, $\preceq$ be a term order on T^X, $M = \{1, t_1, \ldots, t_s\} \subset \mathrm{T}^X$, $M^* = M \setminus \{1\}$. If $x \succ y$ holds for every $x \in X$ and every $y \in Y$, we denote it by $X \succ Y$.

Definition 2.1 (Substitution Set). The set $E = \{y_1 - t_1, \ldots, y_s - t_s\} \subset \mathbb{K}[X \cup Y]$ is called a and the substitution set, where $t_1, \ldots, t_s \in \mathrm{T}^X \setminus \{1\}$. Let $E = 0$ stand for the set $\{y_1 = t_1, \ldots, y_s = t_s\}$.

Definition 2.2 (Extended Ring). $\mathbb{K}[X \cup Y]$ is called an extended ring of $\mathbb{K}[X]$ w.r.t. Y, $I^E = \langle I \cup E \rangle$ is called an extended ideal of I w.r.t. E in $\mathbb{K}[X \cup Y]$, and $\mathrm{T}^{X \cup Y}$ is called an extended term set of T^X w.r.t. Y.

Definition 2.3 (Extended Order). Specify a term order $\preceq$ on T^X. If another term order $\preceq^e$ on $\mathrm{T}^{X \cup Y}$ coincides with $\preceq$ on T^X, then $\preceq^e$ is called an extended order of $\preceq$ on $\mathrm{T}^{X \cup Y}$.

We construct an order on $(\mathrm{T}^X)^N$ based upon the term order $\preceq$ on T^X to describe the relation between the monomial basis of $\mathbb{K}[X]/I$ and the Gröbner basis of I w.r.t. $\preceq$ (cf. Proposition 3.1).

Definition 2.4 (Direct Product). Specify an ordered set $(\mathrm{T}^X, \preceq)$, and two tuples $a = (a_1, \ldots, a_N)$, $b = (b_1, \ldots, b_N) \in (\mathrm{T}^X)^N$, where $\preceq$ is a term order on T^X, N is a positive integer and $(\mathrm{T}^X)^N$ is the Cartesian product. Construct a new ordered set $((\mathrm{T}^X)^N, \preceq')$, where $\preceq'$ is defined by $a \prec' b$ iff $a_i \preceq b_i$ for every $i = 1, \ldots, N$ and there exists an integer j $(1 \leq j \leq N)$ such that $a_j \prec b_j$. Then we call $((\mathrm{T}^X)^N, \preceq')$ the N^{th} direct product of $(\mathrm{T}^X, \preceq)$. (see p. 163, [5])

Remark 2.5. “Direct Product” here is modified slightly, taking place “quasi-order” by “term order”.

Proposition 2.6. $\preceq'$ is a partial order on $(\mathrm{T}^X)^N$.

Proof. First, the reflexivity holds since $a_i \preceq a_i$ $(i = 1,\dots,N)$ for every $a \in (\mathrm{T}^X)^N$. Second, if $a \preceq' b$ and $b \preceq' a$, then for every $i = 1,\dots,N$ we have $a_i \preceq b_i$ and $b_i \preceq a_i$. Hence $a_i = b_i$ and $a = b$, i.e., $\preceq'$ is antisymmetric. Now we only need to verify the transitive property. If $a \preceq' b$ and $b \preceq' c$, then by definition $a_i \preceq b_i$ and $b_i \preceq c_i$ hold for every $i = 1,\dots,N$. Therefore, $a_i \preceq c_i$ and $a \preceq' c$. □

However, it is easy to see that the N^{th} direct order $\preceq'$ in definition 2.4 is not a linear order on $(\mathrm{T}^X)^N$ when $N > 1$.

In the following, we will consider monomial bases of some quotient ring $\mathbb{K}[X]/I$ which are closed under division. While there is no universal name for this concept we will say that it is a *NormalSet* to be in accordance with the name of the corresponding function [10] into the Maple Computer Algebra system Maple: `Groebner[NormalSet]`.

Definition 2.7 (Closed Set). A subset C of T^X is closed iff the two conditions of $m \in C$ and $m^*|m$ imply $m^* \in C$. (see [14] and [36] p. 55)

Definition 2.8 (Normal Set - Closed monomial basis). Consider a zero-dimensional polynomial ideal $I \subset \mathbb{K}[X]$. A closed set $NS \subset \mathrm{T}^X$ is a normal set of I iff the elements of NS form a monomial basis of $\mathbb{K}[X]/I$. (see p. 56, [36])

Here comes the definition of the TSV strategy, the main tool of this paper, where the notions of extended objects play an important role.

Definition 2.9 (TSV strategy). Specify a polynomial ideal $I \subset \mathbb{K}[X]$, a substitution set $E \subset \mathbb{K}[X \cup Y]$ and a term order $\preceq$ on T^X. Compute the Gröbner basis G^E and the corresponding normal set N^E of $I^E = \langle I \cup E \rangle \subset \mathbb{K}[X \cup Y]$ w.r.t. some extended order $\preceq^e$ of $\preceq$ on $\mathrm{T}^{X \cup Y}$; then substitute $E = 0$ into N^E according to Definition 2.1, and a new term set $N \subset \mathrm{T}^X$ is obtained. This strategy to compute G^E, N^E, N or other objects is called the TSV strategy.

Example 2.10. Recall the toy example in the introduction. When TSV strategy is used to repair the artificial discontinuity, in fact we set E to be the binomial set $\{z - xy\}$ and $\preceq^e$ the DRL order with $z \prec y \prec x$. Then the G^E is just the final Gröbner basis G, N^E is $\{1, z, y, x\}$ and N is $\{1, y, x, xy\}$.

Remark 2.11. There are some differences in using the TSV strategy between parametric cases and numerical cases. For the former, we can just use Definition 2.9 to locally repair artificial discontinuities. For the latter, we must consider how to balance the coefficients of the polynomials arising during the Gröbner basis computation, such that the coefficients would not become too small and the loss of accuracy would not be too large. Hence, in numerical cases, we usually add binomials like $y - dt$ to E, where y is a new variable, $|d| \in [\varepsilon, 1]$ and t is a term in the original term set. Nevertheless, the two versions of TSV strategy are *equivalent in theory*.

In the following two sections, we use Definition 2.9 to set up the theory and study the properties.

3. Monomial Bases and Main Theorem

This section is about the main theorem (Theorem 3.4) saying that any monomial basis (containing 1) of $\mathbb{K}[X]/I$ can be obtained with the TSV strategy. When detecting an ε-discontinuity, we add one new variable y and the corresponding binomial $y-t$ into the system and set $y \prec X$. Then t will not arise as a leading term of any polynomial in the Gröbner basis generally. TSV strategy provides us a convenient way in finding other types of monomial basis of the quotient ring.

In order to prove the main theorem, we first give a proposition relating the $\preceq'$-minimal monomial basis of $\mathbb{K}[X]/I$ with the Gröbner basis of I w.r.t. $\preceq$. Papers [14] and [19] sparkled us to study this relation. We will use it to identify the normal set corresponding to the Gröbner basis w.r.t. some term order.

Given a term order $\preceq$ and a term sequence $m_1 \prec m_2 \prec \cdots \in \mathrm{T}^X$ (not necessarily finite), if $\{m_1, m_2, \ldots\}$ composes a basis of $\mathbb{K}[X]/I$, then $m = (m_1, m_2, \ldots)$ is called a *basis tuple* of $\mathbb{K}[X]/I$ in T^X. Denote by MT^X the set of basis tuples of $\mathbb{K}[X]/I$ in T^X.

Proposition 3.1. For a zero-dimensional ideal I, MT^X has a unique element m' such that $m' \preceq' m$ (cf. Definition 2.4) for every $m \in \mathrm{MT}^X$. More precisely, the components of m' form a normal set corresponding to the Gröbner basis G of I w.r.t. $\preceq$.

Proof. Let $m' \in \mathrm{MT}^X$ be the basis tuple formed by the normal set (cf. Definition 2.8) corresponding to the Gröbner basis G of I w.r.t. $\preceq$, such that $1 = m'_1 \prec m'_2 \prec \cdots$, and let m be an element of MT^X. We prove the proposition by induction on i.

- It is easy to see that $1 = m'_1 \preceq m_1$.
- If $m'_k \preceq m_k$ for every $k \leq i$, then $m'_{i+1} \preceq m_{i+1}$.

Otherwise, $m'_{i+1} \succ m_{i+1}$. Note that for every $j \leq i+1$, each term of the normal form of m_j modulo G w.r.t. $\preceq$ (denoted by $\mathrm{NF}_{G,\preceq}(m_j)$) is $\preceq$-smaller than m_j. Hence, $\mathrm{NF}_{G,\preceq}(m_j)$ is a linear combination of $m'_1, \ldots, m'_i$. Therefore, we can see that $m_1, \ldots, m_{i+1}$ are linearly dependent, which is a contradiction.

Uniqueness is obvious. □

Remark 3.2. In fact, the Buchberger-Möller algorithm [22] and the FGLM algorithm [14] are just the processes to find $\preceq'$-minimal elements of their MT^X.

Remark 3.3. When the ideal I in Proposition 3.1 is not zero-dimensional, the result still holds. But the proof and some definitions should be modified in order to deal with infinitely many terms in monomial bases of the quotient ring. We omit the discussion here, since zero-dimensional case is enough in this paper.

Now, we use Proposition 3.1 to prove the main theorem which implies that for a given zero-dimensional polynomial ideal, we can find any monomial basis (containing 1) of the quotient ring with the TSV strategy.

Theorem 3.4. Specify a degree term order $\preceq$ on T^X. A finite term set $N \subset \mathrm{T}^X$ (containing 1) forms a basis of $\mathbb{K}[X]/I$ iff there exists a set of new variables Y such that N can be computed by using the TSV strategy w.r.t. some degree extended term order $\preceq^e$ of $\preceq$ on some extended term set of T^X.

Proof. “$\Leftarrow$” For a polynomial $f(x_1,\dots,x_n,y_1,\dots,y_s) \in \mathbb{K}[X \cup Y]$, a function $\phi(f)$ is defined by $f(x_1,\dots,x_n,t_1,\dots,t_s)+I \in \mathbb{K}[X]/I$. Then, it is easy to see that ϕ is surjective homomorphism and $\mathrm{Ker}(\phi) = I^E$. Therefore, $\mathbb{K}[X \cup Y]/I^E \cong \mathbb{K}[X]/I$. Let

$$\begin{aligned}\psi : \mathbb{K}[X \cup Y]/I^E &\longrightarrow \mathbb{K}[X]/I,\\ h+I^E &\longmapsto \phi(h)\end{aligned}$$

be the isomorphism, where h is a polynomial in $\mathbb{K}[X \cup Y]$. If N^E is the normal set corresponding to the Gröbner basis of I^E w.r.t. some term order on $\mathrm{T}^{X\cup Y}$, then $\phi(N^E)$ is a basis of $\mathbb{K}[X]/I$, i.e., N contains 1 and forms a basis of $\mathbb{K}[X]/I$.

“$\Rightarrow$” Pick $s = \#N-1$. Let $Y = \{y_i : i = 1,\dots,s\}$ be a set of new variables, $E = \{y_i - t_i : y_i \in Y, t_i \in N \setminus \{1\}, i = 1,\dots,s\}$ and $\preceq^e$ be a degree extended term order of $\preceq$ on $\mathrm{T}^{X\cup Y}$. Since ψ is an isomorphism and N is a basis of $\mathbb{K}[X]/I$, the new term set $N' = Y \cup \{1\}$ forms a basis of $\mathbb{K}[X \cup Y]/I^E$. By Proposition 3.1 and Definition 2.3, we know that N' is the normal set corresponding to the Gröbner basis w.r.t. $\preceq^e$. Therefore, N can be computed using the TSV strategy w.r.t. $\preceq^e$. □

In practice, we need not add as many new variables or binomials as in the proof of Theorem 3.4 to the original system. Recall the toy example in the introduction. Though the normal set contains four elements, we only need to add one new variable and one corresponding binomial to locally repair the artificial discontinuity. See Section 6 for more numerical examples (cf. Table 3).

Corollary 3.5. Specify a degree extended term order $\preceq^e$ of $\preceq$ on $\mathrm{T}^{X\cup Y}$ and a finite term set $M \subset \mathrm{T}^X$. If there exists a subset K of $M^* \cup X$ (or M^*) such that $K \cup \{1\}$ forms a basis of $\mathbb{K}[X]/I$, then the set N computed using the TSV strategy w.r.t. $\preceq^e$ is a monomial basis of $\mathbb{K}[X]/I$, more precisely, it is a subset of $M \cup X$ (or M).

Proof. Let N be the term set mentioned above. By Theorem 3.4, N is a monomial basis of $\mathbb{K}[X]/I$. Hence, we only need to prove $N \subset M \cup X$ (or M). Since $K \cup \{1\}$ forms a basis of $\mathbb{K}[X]/I$ and $K \subset M^* \cup X$ (or M^*), there exists a subset X_1 of X and a subset Y_1 of Y such that $X_1 \cup Y_1 \cup \{1\}$ (or $Y_1 \cup \{1\}$) forms a basis of $\mathbb{K}[X \cup Y]/I^E$. By Proposition 3.1 and Definition 2.3, we can know that N^E must correspond to a tuple $m \in \mathrm{MT}^{X\cup Y}$ such that $m \preceq' m'$ where m' corresponds to $X_1 \cup Y_1 \cup \{1\}$ (or $Y_1 \cup \{1\}$). Hence, m must consist of 1 and some elements of $X \cup Y$ (or Y). Therefore, N is a subset of $M \cup X$ (or M). □

From Corollary 3.5, we can see how the TSV computation travels among the monomial bases of $\mathbb{K}[X]/I$ after $M \cup X$ (or M) begins to contain a basis of $\mathbb{K}[X]/I$.

4. Other Relations between Original and Extended Ideals

Since I^E (cf. Definition 2.2) is obtained from I by adding binomials with the new variables linear, it is easy to show that the ideal membership problem is equivalent in I^E and in I; moreover, when I is a zero-dimensional ideal then the number of solutions of I^E or I counting multiplicities are the same and the roots are essentially the same (cf. Proposition 4.2). From that we get the conclusion that I^E and I are equivalent in this sense. Hence, the TSV strategy can also be used to study the two aspects of I in I^E.

The projection of the Gröbner basis G^E obtained by the TSV strategy is generally not a Gröbner basis of the original ideal. The following proposition shows that we can use G^E directly to solve the ideal membership problem just as we use Gröbner bases of the original ideal.

Proposition 4.1. Specify an ideal $I \subset \mathbb{K}[X]$ and a substitution set E. For any $f \in \mathbb{K}[X]$, we have $f \in I$ iff $f \in I^E$.

Proof. It is obvious that if $f \in I$ then $f \in I^E$. Conversely, denote by G_{plex} the reduced Gröbner basis of I^E w.r.t. plex order (Pure Lexicographical order) with $Y \succ X$. Then G_{plex} consists of E and the reduced Gröbner basis of I w.r.t. plex order on T^X. Therefore, if $f \in I^E \cap \mathbb{K}[X]$, then $f \in I$. □

Next, we present a proposition (probably it has been known in other forms) about the relation between the zeros of the original ideal and the extended ideal, which shows that to compute the zeros of I, we can compute them directly in I^E instead of projecting I^E to I.

Proposition 4.2. Specify a zero-dimensional polynomial ideal $I \subset \mathbb{K}[X]$. For any substitution set E, there exists a one-to-one correspondence between the zeros of I in $\overline{\mathbb{K}}^n$ and the zeros of I^E in $\overline{\mathbb{K}}^{n+\#E}$ with the same multiplicities, where $\overline{\mathbb{K}}$ is the algebraic closure of $\mathbb{K}$.

Proof. It is easy to see that I and I' have the same number of distinct zeros $p_1,\ldots,p_r$ and $p'_1,\ldots,p'_r$, where p_k and p'_k have the same first n components in $\mathbb{K}^n$. In what follows, we prove the corresponding zeros of I and I' have the same "intersection multiplicity" (cf. [11] page 139).

Let $\mathscr{O}_k$ be the ring of rational functions defined at p_i, i.e. $\mathscr{O}_k = \{h/g : g(p_i) \neq 0\}$. Let $I\mathscr{O}_k$ be the ideal generated by I in $\mathscr{O}_k$. Let $\phi_k : \mathscr{O}'_k \longrightarrow \mathscr{O}_k/I\mathscr{O}_k$ be a homomorphism such that $\phi_k(h/g) = h(X,M)/g(X,M) + I\mathscr{O}_k$ where $h \in \mathbb{K}[X \cup Y]$ and $g \in \mathbb{K}[X \cup Y] \setminus \langle X - X_0, Y - M_0\rangle$. For any $q \in \mathscr{O}'_k$, if $\phi_k(q) = q(X,M) + I\mathscr{O}_k = I\mathscr{O}_k$, then $q(X,M) = h(X,M)/g(X,M) \in I\mathscr{O}_k$. Hence $h(X,M) \in I\mathscr{O}_k \subset I'\mathscr{O}'_k$. Note that each term in $T^{X\cup Y}$ can be rewritten as follows,

$$\begin{aligned} & x_1^{u_1}\cdots x_n^{u_n} y_1^{v_1}\cdots y_s^{v_s} \\ = \; & x_1^{u_1}\cdots x_n^{u_n}(t_1 + (y_1 - t_1))^{v_1}\cdots(t_s + (y_s - t_s))^{v_s} \\ = \; & x_1^{u_1}\cdots x_n^{u_n} t_1^{v_1}\cdots t_s^{v_s} + \sum_{j=1}^{s} r_j(y_j - t_j) \end{aligned}$$

where $r_j \in \mathbb{K}[X \cup Y]$. Hence,

$$h(X,Y) = h(X,M) + \sum_{j=1}^{s} r_j^*(y_j - t_j) \in I'\mathscr{O}'_k$$

where $r_j^* \in \mathbb{K}[X \cup Y]$. Therefore, $q = h/g \in I'\mathscr{O}'_k$ and $\mathrm{Ker}(\phi_k) \subset I'\mathscr{O}'_k$. Conversely, for every $q \in I'\mathscr{O}'_k$ there exists a $g \in \mathbb{K}[X \cup Y] \setminus \langle X - X_0, Y - M_0\rangle$ with $gq \in I'$. Hence $g(X_0,M_0) \neq 0$ and $(gq)(X_0,M_0) \in I$. Therefore, $\phi_k(q) = I\mathscr{O}_k$ and $\mathrm{Ker}(\phi_k) \supset I'\mathscr{O}'_k$. We have $\mathrm{Ker}(\phi_k) =$

$I'\mathscr{O}'_k$, and $\mathscr{O}'_k/I'\mathscr{O}'_k \cong \mathscr{O}_k/I\mathscr{O}_k$. Consequently, $\dim(\mathscr{O}'_k/I'\mathscr{O}'_k) = \dim(\mathscr{O}_k/I\mathscr{O}_k)$, i.e., p_k and p'_k have the same local intersection multiplicity. □

Remark 4.3. Proposition 4.2 can also be proved easily using the RUR (Rational Univariate Representation) method [27, 28]. We omit the proof here.

5. Algorithms

In numerical computation of Gröbner bases, polynomials with small leading coefficients are not suitable to construct S-polynomials or to reduce other polynomials, because they can make the computation unstable or unreliable [37]. Moreover, we do not know whether they come from the loss of accuracy in most cases, since exact rational computation can also yield very small leading coefficients after the obtained polynomials have been unitized Therefore, we can not consider a coefficient as zero when we find it quite small. So, we would better leave it and do not make such a decision. Now, we clarify what "unitization", " small " and "artificial ε-discontinuity" mean.

Definition 5.1 (Unitization). For any polynomial f, the unitization of f is defined by $\mathscr{U}(f) = f/|c|_{\max}(f)$ where $|c|_{\max}(f) := \max\{|c| : c \text{ is a coefficient of } f\}$. If $f = \mathscr{U}(f)$, we say that f is unitized.

Definition 5.2 (Small). We fix a positive number ε. Given two numbers u and v in $\mathbb{K}$, $|u|$ is said to be *much smaller* than $|v|$ if $|u| < \varepsilon|v|$, denoted by $|u| \ll |v|$. Especially, if $|u| \ll 1$ then we say that u is *small*. For any given polynomial f, we say that the leading coefficient of f is small if $\mathrm{lc}(\mathscr{U}(f))$ is small.

Definition 5.3 (Artificial ε-discontinuity). If the Gröbner basis computation can not continue unless polynomials with small leading coefficients are used to construct S-polynomials or to reduce other polynomials, we say that the polynomial system is an *ε-discontinuous case*. Otherwise, we call it an *ε-continuous case*. For an ε-discontinuous case, if we can compute a monomial basis M (containing 1) of the quotient ring $\mathbb{K}[X]/I$ such that all the polynomials used to construct S-polynomials and to reduce other polynomials are not with small leading coefficients, we say that the system is an *artificial ε-discontinuous case*.

Remark 5.4. In an artificial ε-discontinuous case, we would probably have obtained another monomial basis of $\mathbb{K}[X]/I$ rather than the normal set corresponding to the Gröbner basis.

In this section, we first give two algorithms ε-Buchberger and ε-MatrixF5 to go as far as possible in the Gröbner basis computation. MatrixF5 is an efficient algorithm produced by the first author of this paper, and is a downgraded version of a more efficient algorithm F5 [13]. If the input homogeneous polynomial sequence is "regular", the F5 criterion can avoid reductions to zero in algorithms F5 and MatrixF5. This property is very good for numerical computation of Gröbner bases, since it is an issue to recognize zero polynomial during the numerical reduction. The "ε-" versions of the two existing algorithms mean that they have been modified such that they can avoid small leading coefficients as many as possible. If they stop without outputting the finial Gröbner basis, then we detect an ε-discontinuous case. We also give two algorithms TSVn and TSVh to repair artificial ε-discontinuous cases found by ε-Buchberger and ε-MatrixF5 respectively. Experimental results are shown in the next section.

The ε-Buchberger, based on Buchbeger's algorithm [6, 8, 5], is constructed to compute Gröbner bases. At the same time, we use it to classify the useful polynomials (the polynomials should not be considered as zero) arising in the computation of Gröbner bases into two sets A and B. The polynomials in A are those suitable to construct S-polynomials, and B is composed by those polynomials having small leading coefficients which may, however, be used later. If $|c|_{\max}(f) \not\ll 1$, we denote

$$\mathrm{lt}_{\ll}(f) = \max_{\preceq}\{t : t \in \mathrm{T}(f), |c_t(f)|/|c|_{\max}(f) \not\ll 1\},$$

where $\mathrm{T}(f)$ is the set of terms with nonzero coefficients in f and $c_t(f)$ is the coefficient of term t in f. If $|c|_{\max}(f) \ll 1$, we define $\mathrm{lt}_{\ll}(f) = 0$, and we always consider such polynomials as zero during the computation.

In ε-Buchberger, we use the "normal" selection strategy (first introduced by Buchberger [7]) to select critical pairs. In this strategy, we can only select the critical pair (g_1, g_2) with the minimal $\mathrm{lcm}(\mathrm{lt}(g_1), \mathrm{lt}(g_2))$ w.r.t. a given term order each time. This strategy is often the default selection strategy in Buchberger's algorithm.

Algorithm *ε-Buchberger*

Input: two finite unitized polynomial sets $A, B \subset \mathbb{K}[X]$, a term order $\preceq$ on T^X, a set P of critical pairs and a positive number ε.

Output: two new unitized polynomial sets $A, B \subset \mathbb{K}[X]$

```
$R := \{r : |c|_{\max}(r) \not\ll 1, r = \mathrm{Reduce}(f, A, \preceq), f \in B\}$
$B := \{f : |c|_{\max}(r) \not\ll 1, r = \mathrm{Reduce}(f, A, \preceq), f \in B\}$
if $1 \in \mathrm{lt}_{\ll}(R)$
   then return $[\{1\}, \emptyset, \emptyset]$
while $P \neq \emptyset$ or $R \neq B$
      do while $P \neq \emptyset$
            do select $\{g_1, g_2\} \in P$ by applying the normal strategy
               $P := P \setminus \{g_1, g_2\}$
               $h := \mathrm{spoly}(g_1, g_2, \preceq)$
               $h_0 := \mathrm{Reduce}(h, A, \preceq)$
               if $\mathrm{lt}_{\ll}(h_0) = 1$
                  then return $[\{1\}, \emptyset, \emptyset]$
               if $\mathrm{lt}(h_0) = \mathrm{lt}_{\ll}(h_0) \succ 1$
                  then $P := P \cup \{\{g, \mathscr{U}(h_0)\} : g \in A\}$
                       $A := A \cup \{\mathscr{U}(h_0)\}$
               if $\mathrm{lt}(h_0) \succ \mathrm{lt}_{\ll}(h_0) \succ 1$
                  then $B := B \cup \{\mathscr{U}(h_0)\}$
         $R := \{r : |c|_{\max}(r) \not\ll 1, r = \mathrm{Reduce}(f, A, \preceq), f \in B\}$
         $B := \{f : |c|_{\max}(r) \not\ll 1, r = \mathrm{Reduce}(f, A, \preceq), f \in B\}$
         if $1 \in \mathrm{lt}_{\ll}(R)$
```

```
21.             then return $[\{1\},\emptyset,\emptyset]$
22.         while $R \neq B$
23.             do $W := \{\mathscr{U}(f) : f \in R, \mathrm{lt}(f) = \mathrm{lt}_{\ll}(f) \succ 1\}$
24.                 $A := A \cup W$
25.                 $B := \{\mathscr{U}(f) : f \in R, \mathrm{lt}(f) \succ \mathrm{lt}_{\ll}(f) \succ 1\}$
26.                 $P := \{\{f,h\} : f \in A \setminus \{h\}, h \in W\}$
27.                 $R := \{r : |c|_{\max}(r) \not\ll 1, r = \mathrm{Reduce}(f, A, \preceq), f \in B\}$
28.                 $B := \{f : |c|_{\max}(r) \not\ll 1, r = \mathrm{Reduce}(f, A, \preceq), f \in B\}$
29.                 if $1 \in \mathrm{lt}_{\ll}(R)$
30.                     then return $[\{1\},\emptyset,\emptyset]$
31. return $[A,B]$
```

We can use ε-Buchberger to do both exact and numerical computations. In this algorithm, the function *spoly* is defined as follows:

$$\mathrm{spoly}(g_1,g_2,\preceq) = \frac{S}{\mathrm{lc}_{\preceq}(g_1)} \cdot \frac{T}{\mathrm{lt}_{\preceq}(g_1)} g_1 - \frac{S}{\mathrm{lc}_{\preceq}(g_2)} \cdot \frac{T}{\mathrm{lt}_{\preceq}(g_2)} g_2,$$

where $S = \max\{|\mathrm{lc}_{\preceq}(g_1)|, |\mathrm{lc}_{\preceq}(g_2)|\}$ and $T = \mathrm{lcm}(\mathrm{lt}_{\preceq}(g_1), \mathrm{lt}_{\preceq}(g_2))$.

If we do exact computation, the cycles at lines 6 and 22 terminate, and the termination proof of line 5 is similar with that of Buchberger's algorithm (see p. 214, [5]). Hence, the algorithm ε-Buchberger terminates.

If we do numerical computation, we should consider the loss of accuracy, especially in the functions *Reduce* and *spoly*. If we use a polynomial f with small leading coefficient to reduce or construct S-polynomial with another polynomial g, then in the result we may lose a lot of information of f (cf. [35]). In ε-Buchberger, the leading coefficients of the polynomials in A are not small, so they are suitable to reduce or construct S-polynomials with other polynomials, thus the results of the functions *Reduce* and *spoly* are reliable.

Algorithm *TSVn*

Input: a term order $\mathrm{DRL}(X)$ on T^X, a finite polynomial set $F \subset \mathbb{K}[X]$ with $\langle F\rangle$ zero-dimensional, a positive integer L and a positive number ε

Output: a new variable set Y, a substitution set E, the reduced Gröbner basis G^E of $\langle E \cup F\rangle$ w.r.t. $\mathrm{DRL}(X \cup Y)$

```
1.  $\preceq := \mathrm{DRL}(X)$; $i := 0$; $Y := \emptyset$; $E := \emptyset$
2.  $A := \{\mathscr{U}(f) : f \in F, \mathrm{lt}(\mathscr{U}(f)) = \mathrm{lt}_{\ll}(\mathscr{U}(f)) \succ 1\}$
3.  $B := \{\mathscr{U}(f) : f \in F, \mathrm{lt}(\mathscr{U}(f)) \succ \mathrm{lt}_{\ll}(\mathscr{U}(f)) \succ 1\}$
4.  $P := \{\{g_1,g_2\} : g_1, g_2 \in A, g_1 \neq g_2\}$
5.  $[A,B] := \varepsilon$-Buchberger$(\preceq, A, B, P)$
6.  while $B \neq \emptyset$ or $i < L$
7.      do $i := i+1$
8.          select proper $t_i$ and $d_i$ with $|d_i| \in [\varepsilon, 1]$
9.          $Y := Y \cup \{y_i\}$
10.         $E := E \cup \{d_i t_i - y_i\}$
11.         $P := \{\{g, d_i t_i - y_i\} : g \in A\}$
12.         $A := A \cup \{d_i t_i - y_i\}$
13.         $\preceq := \mathrm{DRL}(X \cup Y)$
```

```
14.         $[A,B]$ :=$\varepsilon$-Buchberger$(\preceq, A, B, P)$
15. **if** $B \neq \emptyset$
16.   **then** print "Can not be repaired in $L$ steps with $E$."
17.         **return**
18. $G^E$ := InterReduce$(A, \preceq)$
19. **return** $[Y, E, G^E]$
```

In algorithm TSVn, t_i and d_i can be chosen to optimize the computation. We use t_i to pivot the leading monomials and use d_i to balance the coefficients arising in the following computation. If we choose only variable as t_i and take $d_i = 1$ each time, the computation is equivalent with changing variable order of the original ring. The positive number L is needed to ensure that TSVn terminate in finite steps, because we do not know whether the input system is an ε-continuous case or an artificial ε-discontinuous case. If the input system is one of these two cases, the G^E obtained is the Gröbner basis of $\langle E \cup F \rangle$ w.r.t. DRL$(X \cup Y)$. Otherwise, the set B would not be empty when the procedure terminates after L steps, and TSVn could not repair the ε-discontinuous case.

We show how TSVn works on a linear example in the next section. In fact, for linear cases, it is just the Gaussian elimination. Note that we do not need to add new variables or binomials to the system, and $B = \emptyset$ in the output of line 5, hence the procedure from line 7 to line 14 need not be called. But for nonlinear cases, ε-Buchberger sometimes can not work well itself, i.e., the output $B \neq \emptyset$, then we must use this part of TSVn.

Before introducing ε-MatrixF5, we must modify Gaussian elimination first, since we need to test whether the leading coefficients of the polynomials are small before they are used to reduce other polynomials. In the following algorithm, we denote the k-th row of $\tilde{M}$ by $\tilde{M}_{k,\cdot}$ and the maximum of absolute values of every entries in row k of $\tilde{M}$ by $\max|\tilde{M}_{k,\cdot}|$.

Algorithm *ε-GaussianElimination*
Input: $M_{m\times n}$ and a positive number ε
Output: j – the column where ε-discontinuity happens, $\tilde{M}$ – simplified matrix

```
1.  $j := 1$, $\tilde{M} := M$, $K := \{1,\dots,m\}$, $D := \emptyset$
2.  **while** $j \leq n$ **and** $K \neq \emptyset$
3.      **do if** $\exists k \in K$ s.t. $\max|\tilde{M}_{k,\cdot}| \not\ll 1$ **and** $|\tilde{M}_{k,j}| \not\ll \max|\tilde{M}_{k,\cdot}|$
4.          **then** take $k^*$ such that $|\tilde{M}_{k^*,j}|/\max|\tilde{M}_{k^*,\cdot}| = \max_{k\in K}\{|\tilde{M}_{k,j}|/\max|\tilde{M}_{k,\cdot}|\}$
5.              $K := K \setminus \{k^*\}$
6.              replace each row $\tilde{M}_{h,\cdot}$ $(h \in K)$ by $\tilde{M}_{h,\cdot} - \tilde{M}_{k^*,\cdot}\tilde{M}_{h,j}/\tilde{M}_{k^*,j}$
7.          **else** **if** $\exists k \in K$ s.t. $\max|\tilde{M}_{k,\cdot}| \ll 1$
8.              **then** $K := K \setminus \{k\}$
9.                  $D := D \cup \{k\}$
10.         **else** **if** $\exists k \in K$ s.t. $\tilde{M}_{k,j} \neq 0$
11.             **then** delete each row $\tilde{M}_{d,\cdot}$ $(d \in D)$ from $\tilde{M}$
12.                 **return** $[j, \tilde{M}]$
13.     $j := j + 1$
14. delete each row $\tilde{M}_{d,\cdot}$ $(d \in D)$ from $\tilde{M}$
15. **return** $[0, \tilde{M}]$
```

If $j = 0$ in the output of ε-GaussianElimination, there is no problem of reduction in the computation and $\tilde{M}$ is the result of Gaussian elimination. Otherwise, j corresponds to a column where the reduction can not continue.

In the following, algorithm ε-MatrixF5 is introduced to deal with homogeneous polynomial systems. However, we can also use it to deal with systems that are not homogenous, since one way to get the Gröbner basis w.r.t. DRL order of an ordinary polynomial ideal is as follows. First homogenize the system with one additional variable, then compute the Gröbner basis G_h w.r.t. DRL order of the homogenized system, and at last dehomogenize G_h (see pp. 466-485, [5]). The algorithm ε-MatrixF5 is derived from Faugère's MatrixF5 which can be considered as a mix algorithm of F4 [12] and F5 [13], and the latter one F5 has the good property that if the input homogeneous polynomial sequence is "regular" then there is no reduction to zero, which is very suitable for numerical computation, since the procedure would not suffer from the identification of 0 polynomial.

Now, we give some descriptions of the notations used in ε-MatrixF5. A row indexed by $s = (i, \tau)$ is used to represent a polynomial obtained as the sum of τf_i and some other "smaller" polynomials in $\langle F \rangle$.

Algorithm *ε-MatrixF5*

Input: X – variable list, F – homogeneous polynomial list with increasing degree,
D – degree bound, ε – positive number

Output: a – leading term where ε-discontinuous case is detected, G' – elements of degree at most D of the Gröbner basis of $\langle F \rangle$ w.r.t. DRL(X)

```
$G' := \emptyset$
for $d$ from $d_{F_1}$ to $D$
    do $\mathscr{M}_{d,0} := \emptyset$, $\tilde{\mathscr{M}}_{d,0} := \emptyset$
        for $i$ from 1 to $\#F$
            do if $d < d_{F_i}$
                then $\mathscr{M}_{d,i} := \tilde{\mathscr{M}}_{d,i-1}$
            if $d = d_{F_i}$
                then $\mathscr{M}_{d,i} :=$ add the new row $F_i$ to $\tilde{\mathscr{M}}_{d,i-1}$ with index $(i, 1)$
            if $d > d_{F_i}$
                then $\mathscr{M}_{d,i} := \tilde{\mathscr{M}}_{d,i-1}$
                    $Crit :=$ LT$(\tilde{\mathscr{M}}_{d-d_{F_i},i-1})$
                    for $s$ in Rows$(\tilde{\mathscr{M}}_{d-1,i})\backslash$Rows$(\tilde{\mathscr{M}}_{d-1,i-1})$
                        do $(i, u) :=$ index$(s)$ with $u = x_{j_1} \cdots x_{j_{d-d_{F_i}-1}}$ and $1 \leq j_1 \leq \cdots \leq j_{d-d_{F_i}-1} \leq \#X$
                            for $j$ from $j_{d-d_{F_i}-1}$ to $\#X$
                                do if $ux_j \notin Crit$
                                    then add the new row $x_j s$ with index $(i, ux_j)$ in $\mathscr{M}_{d,i}$
            $[j_{\text{bad}}, \tilde{\mathscr{M}}_{d,i}] := \varepsilon$-GaussianElimination$(\mathscr{M}_{d,i})$
            if $j_{\text{bad}} \neq 0$
```

```
19.                 then pick the term corresponding to the $j_{\text{bad}}$-th column of $\tilde{\mathscr{M}}_{d,i}$ as
                        $a$
20.                     return $[a, G']$
21.           Add to $G'$ all rows of $\tilde{\mathscr{M}}_{d,\#F}$ not reducible by $\mathrm{LT}(G')$
22.    return $[0, G']$
```

If the input system $\langle F\rangle$ is not an ε-discontinuous case (j_{bad} keeps 0 during the computation), the output G' is a minimal Gröbner basis of $\langle F\rangle$ w.r.t. $\mathrm{DRL}(X)$. Lines 12-16 are based on F5 Criterion [13] to avoid useless computations. If the input polynomial sequence F is regular, then in the ε-GaussianElimination algorithm, lines 7-9 should not be called.

Algorithm *TSVh*
Input: X – variable list, F – homogeneous polynomial list with increasing degree, D – degree band, L – positive integer, ε – positive number
Output: Y – list of new variables, E – homogeneous binomial list, G^E – elements of degree at most D of the reduced Gröbner basis of $\langle F \cup E\rangle$ w.r.t. $\mathrm{DRL}(X \cup Y)$

```
1.  $n := \#X$, $i := 0$, $j := 1$
2.  $Y := \emptyset$, $E := \emptyset$, $G^E := \emptyset$
3.  $[a, G'] := \varepsilon$-MatrixF5$(X, F, D)$
4.  while $a \neq 0$ and $j < L$
5.      do $i := i + 1$
6.          $j := j + 1$
7.          $Y := Y \cup \{y_i\}$
8.          Insert $y_i$ into $X$ between $X_{\#X-1}$ and $X_{\#X} (= x_n)$.
9.          Select proper $t_i$ and $d_i$ with $|d_i| \in [\varepsilon, 1]$ and $\sqrt{\langle t_i\rangle} \neq \langle x_n\rangle$.
10.         $E := E \cup \{d_i t - y_i x_n^{\deg(t)-1}\}$
11.         $F := F \cup \{d_i t - y_i x_n^{\deg(t)-1}\}$
12.         Sort $F$ with increasing degree.
13.         $[a, G'] := \varepsilon$-MatrixF5$(X, F, D)$
14. if $a \neq 0$
15.    then print "Can not be repaired in $L$ steps with $E$"
16.         return
17.    else $G^E := \mathrm{InterReduce}(G')$
18.         return $[Y, E, G^E]$
```

In line 3, we compute the Gröbner basis of the original system $\langle F\rangle$ w.r.t. $\mathrm{DRL}(X)$ by algorithm ε-MatrixF5. If $a \neq 0$, then we need the TSV strategy to pivot the leading term a. In lines 7-13, the TSV strategy is implemented by adding a homogeneous binomial to the system each time, such that it can pivot the leading term a of the corresponding polynomial. In line 13, we call algorithm ε-MatrixF5 again to compute the Gröbner basis of the extended ideal $\langle F\rangle$ w.r.t. $\mathrm{DRL}(X)$. In fact, this process can be greatly optimized since we can reuse the information of $\tilde{\mathscr{M}}_{d,i}$ in the last implementation of ε-MatrixF5. We leave the optimization to the future work.

6. Implementations in Maple and Experiments

In this section, we illustrate how artificial ε-discontinuities can be repaired by the TSV strategy. All the algorithms in this paper were implemented in Maple. We start from a linear case and dense cases, then study some standard benchmarks. From all of these examples, we can see that the TSV strategy is effective on repairing the Gröbner bases for artificial ε-discontinuous cases.

6.1. Linear Case

To see how ε-Buchberger works, we study a linear example first.

Example 6.1 (Linear Case). Consider the following linear system,

$$\begin{cases} f_1 : 0.10519760 \times 10^{-5}x - 0.70383719y - 0.74858720z + 1 \\ f_2 : x - 0.10365584 \times 10^{-5}y + 0.99083786z - 0.74199025 \\ f_3 : 0.12288986x + y - 0.11161051 \times 10^{-5}z - 0.32687369 \\ f_4 : 0.49279128 \times 10^{-1}x + y + 0.63703207z - 0.98207322. \end{cases}$$

Choose $\varepsilon = 10^{-4}$ and a DRL order with $x \succ y$. For a unitized polynomial, if one of its coefficients c satisfies $|c| \leq \varepsilon$, we denote $|c| \ll 1$. Hence, we have $A_1 = \{f_2, f_3, f_4\}$, $B_1 = \{f_1\}$ and $P_1 = \{\{f_2, f_3\}, \{f_2, f_4\}, \{f_3, f_4\}\}$. Then from line 5 of algorithm TSVn we can get $A_2 = \{f_2, f_3, f_4, f_5, f_6, f_7\}$ and $B_2 = \emptyset$, where $f_5 = y - 0.12176504z - 0.23569059$, $f_6 = y + 0.58820438z - 0.94550850$ and $f_7 = z - 0.99978662$. By line 18 of TSVn, we can get $G^E = \{x + 0.24863582, y - 0.35742965, z - 0.99978662\}$. Then, the output of TSVn is $Y = \emptyset$, $E = \emptyset$ and $G^E = \{x + 0.24863582, y - 0.35742965, z - 0.99978662\}$.

6.2. Dense Cases

During the computation of Gröbner basis, if the input system is an ε-continuous case, we only need to apply ε-Buchberger or ε-MatrixF5 without using the TSV strategy. Otherwise, we must use the whole algorithm of TSVn or TSVh. So, a natural question is: How often ε-discontinuous cases can arise when we compute Gröbner bases? We need to have a test. Without loss of generality, we considered random dense systems.

To evaluate the number of ε-discontinuous cases among these systems, we made some testing procedures in Maple. We used random integers as the coefficients of the polynomials, and tested if the leading coefficients of the unitization of these polynomials were smaller than a given band ε. For algorithm ε-Buchberger, we tested random dense zero-dimensional complete intersection quadratic systems. For ε-MatrixF5 algorithm, we studied dense homogeneous quadratic systems. The numbers of ε-discontinuous cases we found are listed in the following table.

When #*polys* ≥ 5, the implementations of ε-Buchberger and ε-MatrixF5 in Maple are slow. So instead, we used the function `fgb_gbasis` in the FGb Maple package [1, 12] to compute the reduced Gröbner basis G of the input system $\langle F \rangle$ w.r.t. DRL order[1]. The difference with ε-Buchberger or ε-MatrixF5 is that it could not stop immediately when the problem happened. We counted the number of systems that have at least one polynomial

[1] This function is available by default to compute Gröbner bases since version 12 of Maple.

TABLE 1. Number of ε-discontinuous cases ($\varepsilon = 10^{-5}$)

#polys	2	3	4	$5^{\dagger}$	$6^{\dagger}$
ε-Buchberger $\times 1000$	0	2	11	61	342
ε-MatrixF5 $\times 1000$	0	0	17	65	334

† tested by the function fgb_gbasis in FGb

$g \in G$ with small leading coefficient. It is easy to see that this number should be no larger than the number of ε-discontinuous cases that should be found by ε-Buchberger or ε-MatrixF5. We used them to estimate the numbers of ε-discontinuities.

From Table 1, we can see that when ε is fixed and the systems become larger, the number of ε-discontinuous cases for dense systems increases rapidly. There are two ways to solve this problem: to decrease ε or to repair the system.

If we decrease ε, the frequency of ε-discontinuities can decrease. For instance, when we chose $\varepsilon = 10^{-7}$, we found 10 ε-discontinuous cases after we tested 200 dense homogenous random systems for $\#polys = 6$. However, note that for floating point computation, generally, the precision σ of numbers is fixed. To avoid large loss of accuracy, the inequality $\varepsilon > \sigma$ ($\varepsilon \gg \sigma$, in practice) has to be satisfied. Hence, we should not hope that the problem can be completely solved only by decreasing ε.

If we repair the system, the frequency of ε-discontinuities can also decrease. We used algorithm TSVn to repair the ε-discontinuous cases in Table 1 found by testing procedure for ε-Buchberger. The result is listed in the table below.

TABLE 2. Number of repaired systems in Table 1 by TSVn ($\varepsilon = 10^{-5}$)

$\#Y$ # polys	1	2	3	4	5	$6 \sim \infty$
4	11	0	0	0	0	0
5	53	6	2	0	0	0
6	224	98	10	6	2	2

From Table 2, we can see that almost all the ε-discontinuous cases has been repaired by adding at most five new variables to the original systems. For ε-discontinuous cases found by ε-MatrixF5, we can repair them by algorithm TSVh similarly. Therefore, the TSV strategy performs perfectly in repairing ε-discontinuities for dense cases.

Based on the preceding discussions, we prefer combining the two ways. First, choose a proper ε for each input system, then repair the system if we detect it an ε-discontinuity during the computation.

Now, another natural question: How to select ε? We describe an idea below. Given a zero-dimensional system ψ (not necessary to be dense) with $\#X = n$ and $\deg(\psi) = D$, we want to find a function $\varepsilon(n, D, \alpha)$ as an upper bound of ε, where $\alpha \in [0, 1]$. To explain the parameter α, we need an ε-discontinuity testing procedure and a TSV performing procedure first. Besides, we also need a positive integer valued function $\Theta(n, D)$.

Via testing a set of dense random systems Φ with each $\varphi \in \Phi$ satisfying $\#X = n$ and $\deg(\varphi) = D$, the function $\varepsilon(n, D, \alpha)$ at each point can be defined as the value of ε such that $\#\{\varphi : \#Y \geq \Theta(n, D)\}/\#\Phi$ tends to α when $\#\Phi$ tends to infinity. For example: Choose the testing and TSV performing procedures used in Table 2. Let $\Theta(n, D) = n$. We can roughly consider $\varepsilon(6, 2, 0.2\%) \approx 10^{-5}$. A table of $\varepsilon(n, D, \alpha)$ can be made to help choose a suitable ε for any input system, no mater it is dense or not. We leave it as a part of future work.

6.3. Standard Benchmarks

To see the effect of the TSV strategy on repairing artificial ε-discontinuities, we also studied some known examples from demos of Jan Verschelde [2] and Jean-Charles Faugère [1]. Since TSVn and TSVh share the same key idea, we only focus on TSVn in the following discussion.

Example 6.2 (Kasura5). Kasura5 is an ε-continuous case for TSVn ($\varepsilon = 10^{-5}$).

Example 6.3 (Tangents0). Tangents0 is an ε-discontinuous case if we take $\varepsilon = 10^{-5}$. Two unitized polynomials in its Gröbner basis w.r.t. $\mathrm{DRL}(x_0, \ldots, x_5)$ have leading coefficients less than 10^{-5}. Their leading monomials are $[0.7152868405 \times 10^{-6}, x_5^5]$ and $[0.2384492025 \times 10^{-5}, x_3 x_5^3]$. Note that this Gröbner basis can be repaired with TSVn by adding a new variable x_6 and a binomial $10^3 x_6 - x_3 x_5$, where $x_3 x_5$ is the square-free part of $x_3 x_5^3$. Then we can find no $|\mathrm{lc}(g)|$ smaller than 10^{-5}, where g is in the new Gröbner basis. Now, the artificial ε-discontinuity has been repaired.

See the following table for more examples.

TABLE 3. Repair of artificial ε-discontinuous cases by TSVn ($\varepsilon = 10^{-5}$)

Systems	Binomials Added
Kasura5	ε-continuous
Sendra	ε-continuous
BenchmarkD1	ε-continuous
Filter9	ε-continuous
ISSAC97	ε-continuous
Tangents0	$1000x_6 - x_3x_5$
Cassou	$480x_5 - x_3(d)$
Lichtblau†	$8000x_4 - x_2^2(x), 8000x_5 - x_3(y), 10x_6 - x_1^2(t), x_7 - x_2(x)$
Cohn2	$6x_5 - x_4(t), 300x_6 - x_3(z), 10^3x_7 - x_1(x), 50x_8 - x_2(y)$
Cohn3‡	$50000x_5 - x_3(z), 50000x_6 - x_2(y), 8000x_7 - x_1(x), 500x_8 - x_4(t)$

†: Though lichtblau is not zero-dimensional, it can also be repaired.
‡: Since the coefficients varied in a larger range, we chose $\varepsilon = 10^{-7}$.

These experiments on standard benchmarks show that the TSV strategy can make the final Gröbner bases avoid artificial ε-discontinuities effectively.

7. Conclusion and Future Work

We provide a strategy called the TSV strategy. Two algorithms ε-Buchberger and ε-MatrixF5 are given to compute Gröbner bases and at the same time check if the input system is an ε-discontinuous case. Based on the main theorem (Theorem 3.4) that all the monomial bases of the quotient ring for zero-dimensional ideals can be worked out with the TSV strategy, the algorithms TSVn and TSVh are produced to repair artificial ε-discontinuities in Gröbner basis computation. Experiments show that the TSV strategy can make the Gröbner basis more suitable for numerical computation if we add proper binomials to the input polynomial system.

Optimizations of these algorithms and implementations in low level languages are planned in the future.

Acknowledgements

The second author would like to thank his co-supervisor Professor Dongming Wang as well as Professors Hoon Hong, Fabrice Rouillier, Lihong Zhi and all the members of SALSA project team for much help. Especially, the second author wants to thank Doctor Wu Fang for her careful corrections.

This work has been supported partially by Chinese SALSA Project, the ANR-NSFC Project ANR-09-BLAN-0371-01/60911130369 (EXACTA) and the SKLSDE Open Fund BUAA-SKLSDE-09KF-01.

References

[1] http://fgbrs.lip6.fr/jcf/software/.

[2] http://www.math.uic.edu/˜jan/.

[3] J. Abbott, C. Fassino, and M.-L. Torrente. Stable border bases for ideals of points. *Journal of Symbolic Computation*, 43(12):883–894, December 2008.

[4] W. Auzinger and H. Stetter. An elimination algorithm for the computation of all zeros of a system of multivariate polynomial equations. In *Conference in Numerical Analysis*, pages 11–30. Birkhäuser-Verlag, 1988.

[5] T. Becker, H. Kredel, and V. Weispfenning. *Gröbner Bases: A Computational Approach to Commutative Algebra*. Springer-Verlag, London, UK, April 1993.

[6] B. Buchberger. *Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal*. PhD thesis, Innsbruck, 1965.

[7] B. Buchberger. Gröbner-Bases: An Algorithmic Method in Polynomial Ideal Theory. In *Multidimensional Systems Theory - Progress, Directions and Open Problems in Multidimensional Systems*, pages 184–232. Reidel Publishing Company, Dodrecht - Boston - Lancaster, 1985.

[8] B. Buchberger. An Algorithm for Finding the Basis Elements in the Residue Class Ring Modulo a Zero Dimensional Polynomial Ideal. *Journal of Symbolic Computation*, 41(3-4), March 2006.

[9] Y. Chen and X. Meng. Border bases of positive dimensional polynomial ideals. In *SNC '07: Proceedings of the 2007 international workshop on symbolic-numeric computation*, pages 65–71, New York, NY, USA, 2007. ACM.

[10] R. M. Corless. Groebner Bases and Matrix Eigenproblems. *SIGSAM Bulletin (Communications in Computer Algebra)*, 30(4):26–32, December 1996.

[11] D. Cox, J. Little, and D. O'Shea. *Using Algebraic Geometry*. Springer-Verlag, Reading, Massachusetts, 2nd edition, 2005.

[12] J.-C. Faugère. A new efficient algorithm for computing Gröbner basis (F4). *Journal of Pure and Applied Algebra*, 139(1-3):61–88, June 1999.

[13] J.-C. Faugère. A new efficient algorithm for computing Gröbner bases without reduction to zero (F5). In *Proceedings of the 2002 International Symposium on Symbolic and Algebraic Computation ISSAC*, pages 75–83, New York, NY, USA, 2002. ACM Press.

[14] J.-C. Faugère, P. Gianni, D. Lazard, and T. Mora. Efficient Computation of Zero-Dimensional Gröbner Basis by Change of Ordering. *Journal of Symbolic Computation*, 16(4):329–344, October 1993.

[15] J.-C. Faugère and Y. Liang. Numrical computation of Gröbner bases for zero-dimensional polynomial ideals. In *Electronic Proceedings of MACIS 2007*, Paris, December 2007. http://www-spiral.lip6.fr/MACIS2007/Papers/.

[16] J.-C. Faugère and Y. Liang. Artificial discontinuities of single-parametric Gröbner bases. *Journal of Symbolic Computation*, 46(4):459–466, April 2011.

[17] A. Kehrein and M. Kreuzer. Characterizations of border bases. *Journal of Pure and Applied Algebra*, 196:251–270, 2005.

[18] A. Kehrein and M. Kreuzer. Computing border bases. *Journal of Pure and Applied Algebra*, 205(2):279–295, 5 2006.

[19] A. Kehrein, M. Kreuzer, and L. Robbiano. An algebraist's view on border bases. In A. Dickenstein and I. Emiris, editors, *Solving Polynomial Equations: Foundations, Algorithms, and Applications*, Algorithms and Computation in Mathematics, pages 160–202, Heidelberg, 2005. Springer Verlag.

[20] A. Kondratyev. Numerical Computation of Gröbner Bases. Technical report, University of Linz, Austria, March 2004. RISC Report Series.

[21] H. Möller. Systems of Algebraic Equations Solved by Means of Endomorphisms. In *AAECC-10: Proceedings of the 10th International Symposium on Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*, pages 43–56. Springer-Verlag, 1993.

[22] H. Möller and B. Buchberger. The construction of multivariate polynomials with preassigned zeros. In *EUROCAM*, pages 24–31, 1982.

[23] B. Mourrain. A New Criterion for Normal Form Algorithms. In *AAECC-13: Proceedings of the 13th International Symposium on Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*, pages 430–443. Springer-Verlag, 1999.

[24] B. Mourrain and P. Trébuchet. Generalized Normal Forms and Polynomial System Solving. In *International Symposium on Symbolic and Algebraic Computation*, pages 253–260. ACM press, July 2005.

[25] G. Reid, J. Tang, J. Yu, and L. Zhi. Hybrid method for solving new pose estimation equation system . In *Proceedings of the 2004 International Workshop on Computer and Geometric Algebra with Applications*, pages 46–57. Springer Berlin Heidelberg, 2005.

[26] G. Reid, J. Tang, and L. Zhi. A Complete Symoblic-Numeric Linear Method for Camera Pose Determination. In *Proceedings of the 2003 International Symposium on Symbolic and Algebraic Computation*, pages 215–223, Philadelphia, Pennsylvania, USA, 2003. ACM Press.

[27] F. Rouillier. Solving zero-dimensional systems through the rational univariate representation. *J. Applicable Algebra in Engineering, Communication and Computing*, 9:433–461, 1999.

[28] F. Rouillier, M.-F. Roy, and M. S. E. Din. Finding at least one point in each connected component of a real algebraic set defined by a single equation. *Journal of Complexity*, 16(4):716–750, 2000.

[29] T. Sasaki and F. Kako. Computing Floating-point Gröbner Bases Stably. In *SNC '07: Proceedings of the 2007 international workshop on symbolic-numeric computation*, pages 180–189, New York, NY, USA, 2007. ACM.

[30] T. Sasaki and F. Kako. Floating-point Gröbner Basis Computation with ill-conditionedness Estimation. In *Proc. of ASCM2007 (LNAI 5081)*, pages 278–292. Springer, 2008.

[31] T. Sasaki and F. Kako. A Practical Method for Floating-point Gröbner Basis Computation. In *Proc. of Joint Conf. of ASCM2009 and MACIS2009*, pages 167–176, 2009.

[32] T. Sasaki and F. Kako. Term cancellations in Computing Floating-point Gröbner Bases. In *Proc. of CASC2010 (LNAI 6244)*, pages 220–231. Springer, 2010.

[33] K. Shirayanagi. Floating point Gröbner bases. In *Selected papers presented at the international IMACS symposium on Symbolic computation, new trends and developments*, pages 509–528, Amsterdam, Netherlands, 1996. Elsevier Science Publishers B. V.

[34] K. Shirayanagi and M. Sweelder. Remarks on Automatic Algorithm Stabilization. *Journal of Symbolic Computation*, 26(6):761–765, December 1998.

[35] H. Stetter. Stabilization of Polynomial Systems Solving with Groebner Bases. In *International Symposium on Symbolic and Algebraic Computation*, pages 117–124. ACM press, 1997.

[36] H. Stetter. *Numerical Polynomial Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2004.

[37] C. Traverso and A. Zanoni. Numerical stability and stabilization of Groebner basis computation. In *International Conference on Symbolic and Algebraic Computation*, pages 262–269, New York, NY, USA, 2002. ACM Press.

[38] P. Trébuchet. Generalized normal forms for positive dimensional ideals. In *International Conference on Polynomial System Solving*, 2004.

[39] V. Weispfenning. Gröbner Bases for Inexact Input Data. In *Proc. of CASC2003*, pages 403–411, Passau, Germany, 2003.

Jean-Charles Faugère
INRIA, Paris-Rocquencourt Center, SALSA Project
UPMC, Univ Paris 06, LIP6
CNRS, UMR 7606, LIP6
UFR Ingénierie 919, LIP6, Case 169, 4, Place Jussieu, F-75252 Paris
e-mail: `Jean-Charles.Faugere@inria.fr`

Ye Liang
LMIB, School of Mathematics and Systems Sciences,
Beihang University,
37 Xueyuan Road, Haidian District, 100191 Beijing, CHINA and

INRIA, Paris-Rocquencourt Center, SALSA Project
UPMC, Univ Paris 06, LIP6
CNRS, UMR 7606, LIP6
UFR Ingénierie 919, LIP6, Case 169, 4, Place Jussieu, F-75252 Paris;

KLMM, Institute of Systems Science,
Academy of Mathematics and System Science,
Chinese Academy of Sciences,
55 Zhongguancun East Road, Haidian District, 100190 Beijing, CHINA
e-mail: `wolf39150422@gmail.com`