Formalization of Dubé's Degree Bounds for Gröbner Bases in Isabelle/HOL

Alexander Maletzky[0000−0003−4378−7854],⋆

RISC, Johannes Kepler University Linz, Austria,
alexander.maletzky@risc.jku.at

Abstract. We present an Isabelle/HOL formalization of certain upper bounds on the degrees of Gröbner bases in multivariate polynomial rings over fields, due to Dubé. These bounds are not only of theoretical interest, but can also be used for computing Gröbner bases by row-reducing Macaulay matrices.
The formalization covers the whole theory developed by Dubé for obtaining the bounds, building upon an extensive existing library of multivariate polynomials and Gröbner bases in Isabelle/HOL. To the best of our knowledge, this is the first thorough formalization of degree bounds for Gröbner bases in any proof assistant.

1 Introduction

Gröbner bases [3,1] are one of the most powerful and most widely used tools in modern computer algebra, as they allow to effectively solve many problems related to ideals of multivariate polynomial rings. More precisely, Gröbner bases are generating sets of ideals with certain additional properties; Buchberger, in his doctoral thesis [3], proved that every ideal has a finite Gröbner basis and even provided an algorithm for computing one. Later, Dubé in [4] derived general upper bounds on the degrees of polynomials in Gröbner bases, which only depend on the number of indeterminates and the maximum degree of the polynomials in *some* generating set on the ideal under consideration.

Upper bounds for the degrees of Gröbner bases are not only of theoretical interest, but also have practical relevance. For instance, Wiesinger-Widi in [14] shows that Gröbner bases can be computed by row-reducing a certain *Macaulay matrix* corresponding to the input set. The Macaulay matrix of a set of polynomials is just a huge matrix whose entries are the coefficients of the given polynomials. The caveat of this approach is, however, that upper bounds on the degrees of the resulting Gröbner bases must be known a-priori – and thanks to Dubé such bounds can be easily computed for every input set.

In this paper we describe a recent formalization of Dubé's bounds in the Isabelle/HOL proof assistant [11,13]; to the best of our knowledge, it is the first-ever formal treatment of said bounds in *any* proof assistant. In a nutshell,

⋆ The research was funded by the Austrian Science Fund (FWF): P 29498-N31

the formalization introduces the constant $\text{Dube}_{n,d}$ depending on $n, d \in \mathbb{N}_0$ such that, for every finite set F of multivariate polynomials in n indeterminates having maximum degree d, there exists a Gröbner basis G of F the degrees of whose elements are at most $\text{Dube}_{n+1,d}$ (cf. Corollary 2). Although this statement might look innocent, the machinery needed to prove it is fairly extensive: one needs to define so-called *cone decompositions* of the polynomial ring, establish their relationship to the *Hilbert function*, and finally do a series of algebraic manipulations and estimations involving sums over generalized binomial coefficients. In the upcoming sections we not only show the formalized definitions and theorems in Isabelle/HOL, but also present their informal counterparts and therefore try to make the paper as self-contained as possible; still, due to the lack of space, we cannot include all the details (let alone proofs), so the interested reader is referred to [4] instead.

The formalization is freely available online as a GitHub repository [8, theory 'Dube_Bound'] and is compatible with the development versions of Isabelle[1] and the Archive of Formal Proofs[2]. We plan to submit it to the Archive of Formal Proofs eventually.

The remainder of this paper is organized as follows: In Section 2 we briefly recall the necessary mathematical background of multivariate polynomials and Gröbner bases, and also present the foundations in Isabelle/HOL our formalization builds upon. Section 3 is all about cone decompositions, the key tool to obtaining Dubé's bounds. Section 4, then, briefly sketches how the final bounds can eventually be obtained and also lists the main theorems. Section 5 concludes the paper.

2 Preliminaries

2.1 Mathematical Background

We briefly review the most important mathematical concepts appearing in our formalization. Clearly, since our main goal is to prove degree bounds on Gröbner bases, we have to begin by explaining the rough idea behind Gröbner bases.

First of all, we fix a field K and a finite set of indeterminates $X = \{x_1, \ldots, x_n\}$; $K[X]$, then, denotes the multivariate polynomial ring over K, i. e. the set of all finite sums of the form $c_1 t_1 + \ldots + c_m t_m$, where the c_i are non-zero coefficients in K and the t_i are *power-products* of the form $x_1^{e_1} \cdots x_n^{e_n}$ ($e_j \in \mathbb{N}_0$). The set of all power-products appearing in a polynomial p with non-zero coefficient is called the *support* of p, denoted by $\text{supp}(p)$. The set of *all* power-products over X will be denoted by $[X]$.

We now fix an *admissible order relation* $\preceq$ on $[X]$, that is, $\preceq$ is a well-order on $[X]$ additionally satisfying (i) $1 \preceq t$ for all $t \in [X]$, and (ii) $s \preceq t \Rightarrow us \preceq ut$ for all $s, t, u \in [X]$. Note that there are infinitely many such admissible order relations if $n > 1$.

[1] `http://isabelle.in.tum.de/repos/isabelle`

[2] `http://devel.isa-afp.org/`

With respect to $\preceq$, every polynomial $p \neq 0$ possesses a *leading power-product* $\mathrm{lpp}(p)$: this is the unique largest (w. r. t. $\preceq$) power-product in $\mathrm{supp}(p)$.

Finally, we recall the definition of an *ideal*: in a commutative ring R with unit, and ideal is a subset of R that is closed under addition and under multiplication by arbitrary elements of R. The ideal generated by a set $B \subseteq R$, denoted by $\langle B \rangle$, is the (unique) smallest ideal containing B.

Now, we have all the ingredients for defining Gröbner bases [3,1]: a *Gröbner basis* $G \subseteq K[X] \backslash \{0\}$ is a *finite* set of polynomials such that, for all $p \in \langle G \rangle \backslash \{0\}$, there exists a $g \in G$ such that $\mathrm{lpp}(g) \mid \mathrm{lpp}(p)$ (where '|' is the usual divisibility relation on $[X]$). Gröbner bases have a lot of interesting properties, of which we summarize the most important ones here:

- Every ideal in $K[X]$ has Gröbner basis, i. e. for every $F \subseteq K[X]$ there exists some $G \subseteq K[X]$ that is a Gröbner basis and satisfies $\langle G \rangle = \langle F \rangle$. For the sake of simplicity, G is also called a Gröbner basis of F (not just $\langle F \rangle$).
- By imposing additional constraints on Gröbner bases, one can even make them unique for every ideal. This leads to the concept of *reduced Gröbner bases*: every ideal in $K[X]$ has a unique reduced Gröbner basis.[3]
- If G is a Gröbner basis, then for every $p \in K[X]$ there exists a *unique* $q \in K[X]$ such that (i) $\mathrm{lpp}(g) \nmid t$ for all $t \in \mathrm{supp}(q)$ and $g \in G$, and (ii) $p - q \in \langle G \rangle$. Hence, if $F \subseteq K[X]$ is arbitrary, we can define the *normal form* (w. r. t. F) of any polynomial p to be the unique q satisfying the above two properties for the reduced Gröbner basis of F. We will denote the normal form of p w. r. t. F by $\mathrm{nf}_F(p)$. Note that $\mathrm{nf}_F(p) = 0$ if, and only if, $p \in \langle F \rangle$.
- (Reduced) Gröbner bases and normal forms are effectively computable, but the details are not so important here.

Besides Gröbner bases, we need a couple of other notions related to polynomials which we briefly introduce/recall:

- For $F \subseteq K[X]$, the set $\mathrm{N}_F \subseteq K[X]$ contains precisely the polynomials that are in normal form w. r. t. F, i. e. those p satisfying $\mathrm{nf}_F(p) = p$. Apparently $\langle F \rangle \cap \mathrm{N}_F = \{0\}$.
- For $z \in \mathbb{N}_0$ and $p \in K[X]$, the *homogeneous component* of p at z is the sub-polynomial of p all of whose power-products have degree z. Hence, p can be written as $p = \sum_{z=0}^{\infty} p_z$. If $T \subseteq K[X]$, then T_z is the set of the homogeneous components of the elements of T at z.
- A polynomial is *homogeneous* if, and only if, it has at most one non-zero homogeneous component, i. e. all power-products in its support have the same degree. A set $T \subseteq K[X]$ is called homogeneous if it is a K-vector space and, for every $p \in T$ and $z \in \mathbb{N}_0$, $p_z \in T$ as well. Obviously, if T is homogeneous, then T_z constitutes a K-vector space for all z. Also, it is well-known that an ideal is homogeneous if, and only if, it can be generated by a set of homogeneous polynomials. N_F is always homogeneous.

[3] Reduced Gröbner bases are unique for every admissible order relation, but different orders may yield different reduced Gröbner bases for the same ideal.

– If $T \subseteq K[X]$ is homogeneous, then the *Hilbert function* of T, denoted by $\varphi_T(z)$, maps every $z \in \mathbb{N}_0$ to the dimension of T_z as a K-vector space.
– Let $T \subseteq K[X]$ and $S_1, \ldots, S_m \subseteq T$. Then the S_i form a *direct decomposition* of T, written $T = S_1 \oplus \cdots \oplus S_m$, if every $p \in T$ can be *uniquely* expressed as a sum of the form $p = \sum_{i=1}^{m} s_i$ with $s_i \in S_i$ for all $1 \leq i \leq m$. If furthermore each of the S_i is homogeneous, then $\varphi_T(z) = \sum_{i=1}^{m} \varphi_{S_i}(z)$. Also, it is easy to see that $K[X] = \langle F \rangle \oplus \mathrm{N}_F$ for all $F \subseteq K[X]$, hence

$$\binom{z+n-1}{n-1} = \varphi_{K[X]}(z) = \varphi_{\langle F \rangle}(z) + \varphi_{\mathrm{N}_F}(z) \tag{1}$$

if F contains only homogeneous polynomials.

2.2 Isabelle/HOL

Due to space limitations we have to presuppose from the reader basic knowledge of the (mostly self-explanatory) syntax of Isabelle/HOL. We only slightly deviate from the usual Isabelle/HOL syntax in that the image of a set A under a function f will be denoted by $f \bullet A$ here, to ease readability. So, for instance $(*)\ x \bullet A$ is a compact representation of the set $\{x * a \mid a \in A\}$.

Multivariate polynomials and Gröbner bases have already been formalized in Isabelle/HOL in [12,6]. We only recall the most important aspects of these formalizations here, to make the paper as self-contained as possible. The fundamental concept underlying polynomials are so-called *polynomial mappings*, which are functions from some type α to some other type $\beta :: \texttt{zero}$, such that only finitely many arguments are mapped to non-zero values. The type of such polynomial mappings is $\alpha \Rightarrow_0 \beta$. So, a multivariate polynomial with indeterminates of type χ and coefficients of type α is simply a term of type $(\chi \Rightarrow_0 \texttt{nat}) \Rightarrow_0 \alpha$: $\chi \Rightarrow_0 \texttt{nat}$ is the type of power-products, where every indeterminate is mapped to some exponent, and in terms of type $(\chi \Rightarrow_0 \texttt{nat}) \Rightarrow_0 \alpha$ every power-product is mapped to some coefficient. Throughout the paper, the type $\chi \Rightarrow_0 \texttt{nat}$ is abbreviated by χ `pp`, the type χ `pp` $\Rightarrow_0 \alpha$ is abbreviated by (χ, α) `poly`, the type variable χ always represents the type of indeterminates, and α always represents the type of coefficients (usually tacitly assumed to belong to sort `field`).

For a set X of indeterminates of type χ, $.[X]$ is the formalization of the set of power-products in X (i. e. $[X]$, but in the formal sources we had to add a dot to distinguish it from a singleton list), and $\texttt{P}[X]$ is the formalization of the set of polynomials with power-products in $.[X]$ (roughly corresponds to $K[X]$, but we chose the letter '`P`' because '`K`' would suggest the coefficients be in a field, whereas $\texttt{P}[X]$ does not impose any restrictions on the coefficients). $\texttt{P}[X]$ and $.[X]$ are needed, because often we have to consider subsets of the whole types (χ, α) `poly` and χ `pp`, respectively, where only indeterminates in X occur.

The following are the formalizations of the basic concepts related to polynomials and Gröbner bases that are mentioned in Section 2.1: `reduced_GB` F is the reduced Gröbner basis of F (w. r. t. the admissible order relation $\preceq$ which is implicitly fixed in a locale); `ideal` F corresponds to $\langle F \rangle$; `normal_form` F p

corresponds to $\mathrm{nf}_F(p)$, and hence `normal_form` F • `UNIV` corresponds to N_F (if the indeterminates shall be restricted to X it is `normal_form` F • `P`$[X]$); `deg_pm` t and `poly_deg` p refer to the degree of a power-product and of a polynomial, respectively[4]; `homogeneous` p expresses that p is a homogeneous polynomial; `direct_decomp` T ss states that the list of sets ss constitutes a direct decomposition of the set T; and `Hilbert_fun` T z, finally, is the Hilbert function of T. Since there is nothing interesting about the formal definitions of these concepts in Isabelle/HOL, we omit them here, but interested readers might want to have a look at [9].

Remark 1. This remark is only relevant for readers intending to look at the actual Isabelle sources of the formalization. There, power-products are written *additively* rather than multiplicatively, for technical reasons. This means that 0 is used in the place of 1 and $+$ in the place of $\cdot$. In this paper we use the standard multiplicative notation for the sake of clarity, though.

3 Cone Decompositions

3.1 Basics

The key to obtaining Dubé's degree bounds is decomposing the ring $K[X]$ into so-called *cones*, which are subsets of $K[X]$ whose Hilbert functions can be described easily.

Definition 1. *Let $h \in K[X]\backslash\{0\}$, and let $U \subseteq X$. Then the* cone *of h and U, denoted by* $\mathrm{cone}(h, U)$*, is the set $\{gh \mid g \in K[U]\}$.*

A cone decomposition *of a set $T \subseteq K[X]$ is a finite set $\{(h_1, U_1), \ldots, (h_r, U_r)\}$ of pairs such that $T = \mathrm{cone}(h_1, U_1) \oplus \cdots \oplus \mathrm{cone}(h_r, U_r)$. If the h_i are homogeneous, we call the cone decomposition* homogeneous*, and if the h_i are monic monomials we call the cone decomposition a* monomial cone decomposition*.*

To gain some intuition about cones, the interested reader is referred to [4]. Roughly, a cone $\mathrm{cone}(h, U)$ corresponds to a principal ideal, where multiplication is only allowed by polynomials in $K[U]$ rather than $K[X]$, though. Note, however, that h is still a polynomial in $K[X]$, and may therefore contain indeterminates in $X\backslash U$.

Definition 2. *Let P be a cone decomposition of some set T and $k \in \mathbb{N}_0$. P is called* k-standard *if, and only if, for all $(h, U) \in P$ with $U \neq \emptyset$ it holds that (i) $\deg(h) \geq k$, and (ii) for every $k \leq d \leq \deg(h)$ there exists $(g, V) \in P$ with $\deg(g) = d$ and $|V| \geq |U|$.*

Formalizing the above definitions in Isabelle/HOL is absolutely straightforward, building upon the library of multivariate polynomials sketched in Section 2.2. Before, however, we fix a finite set X of indeterminates (of type χ) in a local theory context:

[4] `poly_deg` 0 is defined to be 0.

```
context
  fixes X :: "χ set"
  assumes "finite X"
begin
```

So, all subsequent definitions and lemmas are implicitly parameterized over the finite set X. The reason for fixing X in this way is that the type χ is not necessarily finite; indeed, it is very convenient to be able to instantiate χ by type `nat` if one wishes to have an infinite supply of indeterminates. However, most results we are going to formalize are only valid in polynomial rings with *finitely* many indeterminates, meaning that very often we have to add explicit assumptions of the form $p \in \mathtt{P}[X]$ or $F \subseteq \mathtt{P}[X]$, as will be seen below.

These are now the definitions of cones and cone decompositions, respectively:

```
definition cone :: "(((χ, α) poly × χ set) ⇒ ((χ, α) poly) set"
  where "cone hU = (∗) (fst hU) ● P[snd hU]"

definition cone_decomp :: "((χ, α) poly) set ⇒
                           ((χ, α) poly × χ set) list ⇒ bool"
  where "cone_decomp T ps = direct_decomp T (map cone ps)"
```

There are a few things to note:

- In the formalization, constant `cone` is defined for pairs of arguments rather than two individual arguments, i. e. it is uncurried. This turned out more convenient for our purpose.
- Also, `cone` (h, U) does not check whether $h \in K[X]\backslash\{0\}$ and whether $U \subseteq X$. Instead, we introduced the predicate `valid_decomp` which, for a given cone decomposition, performs that check on all pairs in it explicitly.
- Cone decompositions are defined for lists of pairs rather than sets of pairs. This is mainly because `direct_decomp` is also defined for lists, and it also allows us to avoid many explicit finiteness checks we would have to make otherwise (which would be feasible but inconvenient).

Besides `cone`, `cone_decomp` and `valid_decomp` there are also `hom_decomp` *ps*, `monomial_decomp` *ps* and `standard_decomp` *k ps*, which express that *ps* is a homogeneous, monomial or k-standard cone decomposition, respectively.

Here comes the first important property of k-standard cone decompositions:

```
lemma standard_decomp_geE:
  assumes "valid_decomp X ps" "cone_decomp T ps" "standard_decomp k ps" "k ≤ d"
  obtains qs where "valid_decomp X qs" "cone_decomp T qs" "standard_decomp d qs"
    "monomial_decomp ps ⟹ monomial_decomp qs" "hom_decomp ps ⟹ hom_decomp qs"
```

This lemma states that for any k-standard cone decomposition P of T, and for all $d \geq k$, there exists a d-standard cone decomposition Q of T; furthermore, if P is a homogeneous or monomial cone decomposition, then so is Q. The proof of *standard-decomp-geE* is based on the fact that for any $h \in K[X]\backslash\{0\}$ and $U = \{x_{i_1}, \ldots, x_{i_m}\} \subseteq X$, the set

$$\{(h, \emptyset)\} \cup \{(x_{i_j} h, \{x_{i_j}, \ldots, x_{i_m}\}) \mid 1 \leq j \leq m\}$$

constitutes a $(\deg(h)+1)$-standard cone decomposition of $\mathrm{cone}(h,U)$.

Remark 2. Apparently, Lemma *standard-decomp-geE* only asserts the *existence* of the desired d-standard cone decomposition Q, but does not provide an algorithm for actually *computing* it – despite the fact that its proof is actually constructive. This reveals a general principle we adhered to throughout the formalization: many lemmas only stipulate the *existence* of certain cone decompositions, without showing how to construct them. The reason for this inconstructive approach is simple: cone decompositions merely constitute a theoretical artifact needed for obtaining the desired degree bounds, but they do not appear in these bounds at all. The final degree bounds *are* effectively computable, of course.

3.2 Cone Decompositions of $\langle F\rangle$ and N_F

Our next goal is to construct cone decompositions with certain special properties of both $\langle F\rangle$ and N_F, where $F \subseteq K[X]$. More precisely, we formally prove the following two theorems:

Theorem 1 (Theorem 4.11. in [4]). *Let $F \subseteq K[X]$. Then there exists a 0-standard cone decomposition Q of N_F. Moreover, if the polynomials in F are homogeneous, then $\deg(g) \leq d$ for all g in the reduced Gröbner basis of F, where $d = 1 + \max\{\deg(h) \mid (h,U) \in Q\}$.*

Theorem 2 (Corollary 5.2. in [4]). *Let $F \subseteq K[X]$ be finite, let $f \in F$, and assume that f has maximal degree among all polynomials in F. Then there exist $T \subseteq K[X]$ and a $\deg(f)$-standard cone decomposition P of T such that $\langle F\rangle = \langle f\rangle \oplus T$. Moreover, if the polynomials in F are homogeneous, then P is a homogeneous cone decomposition.*

As can be seen, Theorem 1 already provides some sort of degree bound for the reduced Gröbner basis of a set F, depending, however, on a cone decomposition of N_F. How it is possible to get rid of that cone decomposition, and what Theorem 2 is needed for, will be illustrated in Sections 3.3 and 4. But first let us have a look at the formalizations of the two theorems in Isabelle/HOL.

As a starting point, we need to introduce the following auxiliary concept:

```
definition splits_wrt ::
             "(((χ, α) poly × χ set) list × ((χ, α) poly × χ set) list) ⇒
                 (χ, α) poly set ⇒ (χ, α) poly set ⇒ bool"
  where "splits_wrt pqs T F = (let ps = fst pqs; qs = snd pqs in
                                cone_decomp T (ps @ qs) ∧
                                (∀hU∈set ps. cone hU ⊆ ideal F) ∧
                                (∀(h, U)∈set qs. cone (h, U) ∩ ideal F = {0}))"
```

Informally, `splits_wrt` (P,Q) T F asserts that (i) $P \cup Q$ is a cone decomposition of T, (ii) $\mathrm{cone}(h,U) \subseteq \langle F\rangle$ for all $(h,U) \in P$, and (iii) $\mathrm{cone}(h,U) \cap \langle F\rangle = \{0\}$ for all $(h,U) \in Q$.[5]

[5] Of course, `splits_wrt` is defined for *lists* ps, qs instead of sets P, Q, but informally it is easier to think of sets.

One can prove that, under the assumption `splits_wrt` (P, Q) T F, P and Q are cone decompositions of certain sets:

```
lemma splits_wrt_cone_decomp_1:
  assumes "splits_wrt (ps, qs) T F" "monomial_decomp qs" "is_monomial_set F"
  shows "cone_decomp (T ∩ ideal F) ps"

lemma splits_wrt_cone_decomp_2:
  assumes "splits_wrt (ps, qs) T F" "monomial_decomp qs" "is_monomial_set F"
          "F ⊆ P[X]"
  shows "cone_decomp (T ∩ normal_form F ` P[X]) qs"
```

This already looks promising, because if T is sufficiently large (e. g. the whole ring $K[X]$), these two lemmas assert that P and Q are cone decompositions of $\langle F \rangle$ and N_F, respectively – at least if F only consists of monomials. So, it would be great if we could prove that (under some restrictions on T and F) there always exist P and Q satisfying `splits_wrt` (P, Q) T F. And indeed, such P and Q *do* exist, because they can be constructed by the recursive function `split`:

```
function split :: "'χ pp ⇒ 'χ set ⇒ 'χ pp set ⇒
        ((('χ, 'α) poly × ('χ set)) list) × ((('χ, 'α) poly × ('χ set)) list))"
  where
    "split t U S =
      (if 1 ∈ S then
        ([(monomial 1 t, U)], [])
      else if S ∩ .[U] = {} then
        ([], [(monomial 1 t, U)])
      else
        let x = (SOME x'. x' ∈ U - (max_subset U (λV. S ∩ .[V] = {})));
            (ps0, qs0) = split t (U - {x}) S;
            (ps1, qs1) = split (single x 1 * t) U ((λs. s / single x 1) ` S) in
          (ps0 @ ps1, qs0 @ qs1))"
```

Some remarks on the above Isabelle/HOL code are in place:

- `split` t U S consists of three branches: the first two branches correspond to the base cases without any recursive calls, where ps and qs (i. e. P and Q) can be determined readily. In the third branch, `split` is applied recursively twice, producing ps_0, qs_0, ps_1 and qs_1, whose concatenations are eventually returned.
- `max_subset` U $(\lambda V.\ S \cap .[V] = \{\})$ returns a maximal $V \subseteq U$ satisfying $S \cap [V] = \emptyset$. x is then chosen as some element in $U \backslash V$, which is not empty by case assumption.
- `monomial` 1 t represents the monomial whose coefficient is 1 and whose sole power-product is t; likewise, `single` x 1 represents the power-product in which the exponent of x is 1 and all other exponents are 0.
- Termination of `split` is not entirely obvious, but under some mild assumptions on its input the function does indeed terminate.
- We did not take the effort to make function `split` executable, because of the arguments put forward in Remark 2. In principle, this would be possible, though.

The fact that the result returned by function `split` really splits a set w. r. t. another set, as claimed above, is stated in the next lemma:

```
lemma split_splits_wrt:
  assumes "U ⊆ X" "finite S" "t ∈ .[X]" "ideal F ÷ t = ideal (monomial 1 ` S)"
  shows "splits_wrt (split t U S) (cone (monomial 1 t, U)) F"
```

So, the set T that is split w. r. t. F is actually the cone $\mathrm{cone}(t, U)$. This suffices for our purpose, because for $t = 1$ and $U = X$ we apparently have $\mathrm{cone}(t, U) = K[X]$, and therefore the output of `split` 1 X F really constitutes a cone decomposition of $\langle F\rangle$ and N_F, respectively, just as desired. In the assumption of Lemma *split-splits-wrt*, $\langle F\rangle \div t$ denotes the *quotient ideal* of $\langle F\rangle$ w. r. t. t, i. e. the ideal $\{p \in K[X] \mid t\,p \in \langle F\rangle\}$. So, that assumption basically demands that $\langle F\rangle \div t$ be generated by a finite set S of monomials.

Before we can prove Theorem 1, we need another crucial property of function `split` that guarantees the existence of cones of a certain shape in its second return value:

```
lemma lem_4_8:
  assumes "finite S" "S ⊆ .[X]" "1 ∉ S" "g ∈ reduced_GB (monomial 1 ` S)"
  obtains t U where "U ⊆ X" "(monomial 1 t, U) ∈ set (snd (split 1 X S))"
    "poly_deg g = deg_pm t + 1"
```

Together with Lemma *split-splits-wrt*, *lem-4-8* is the key to proving Theorem 1, because the desired cone decomposition Q can be shown to be precisely the second return value of `split` when applied to the appropriate input. The formalization of Theorem 1, hence, looks like this:

```
theorem standard_cone_decomp_snd_split:
  assumes "F ⊆ P[X]"
  defines "qs = snd (split 1 X (lpp ` reduced_GB F))"
  defines "d = 1 + Max (poly_deg ` fst ` set qs)"
  shows "standard_decomp 0 qs" "cone_decomp (normal_form F ` P[X]) qs"
    "(∀f∈F. homogeneous f) ⟹ g ∈ reduced_GB F ⟹ poly_deg g ≤ d"
```

This theorem connects the second return value of function `split` to a cone decomposition of N_F, and as the attentive reader can probably guess, the first return value of `split` can be utilized to obtain the desired direct decomposition $\langle F\rangle = \langle f\rangle \oplus T$, where T has a $\deg(f)$-standard cone decomposition P. Unfortunately, however, the first return value of `split` cannot be directly used for that purpose, but has to be slightly adjusted (especially for making it $\deg(f)$-standard). The details are a bit technical and can be found in [4, Section 5]; here, we only show the final formal statement of Theorem 2:

```
theorem ideal_decompE:
  assumes "finite F" "F ⊆ P[X]" "f ∈ F" "∀f'∈F. poly_deg f' ≤ poly_deg f"
  obtains T ps where "valid_decomp ps" "standard_decomp (poly_deg f) ps"
    "cone_decomp T ps" "direct_decomp (ideal F ∩ P[X]) [ideal {f} ∩ P[X], T]"
    "(∀f'∈F. homogeneous f') ⟹ hom_decomp ps"
```

Please note that we cannot simply write `ideal` F and `ideal` $\{f\}$, but we explicitly have to restrict these sets to polynomials in $\mathtt{P}[X]$, since the `ideal` operator allows multiplication by polynomials in arbitrary indeterminates of type χ.

3.3 Exact Cone Decompositions

We begin by summarizing the main definitions and results of this subsection informally:

Definition 3. *Let Q be a cone decomposition. Q is called* exact *if, and only if, for all pairs $(h,U),(g,V) \in Q$, if $U \neq \emptyset$, $V \neq \emptyset$ and $\deg(h) = \deg(g)$, then $(h,U) = (g,V)$.*

Definition 4. *Let Q be a cone decomposition. Then $\mathfrak{a}_Q$ is the smallest $k \in \mathbb{N}_0$ such that Q is k-standard, or 0 if Q is not k-standard for any k.*

Also, the sequence $\mathfrak{b}_{Q,i}$ for $i \in \mathbb{N}_0$ is defined as

$$\mathfrak{b}_{Q,i} := \min\{d \geq \mathfrak{a}_Q \mid \forall (h,U) \in Q : |U| \geq i \Longrightarrow \deg(h) < d\}.$$

Similar to the previous section, the main result of this section ensures the existence of certain cone decompositions, in this case exact ones:

Theorem 3 (Lemma 6.3. in [4]). *Let Q be a k-standard cone decomposition of some set T. Then there also exists an* exact *k-standard cone decomposition Q' of T such that $\max\{\deg(h) \mid (h,U) \in Q\} \leq \max\{\deg(h) \mid (h,U) \in Q'\}$. Furthermore, if Q is a homogeneous or monomial decomposition, then so is Q'.*

Theorems 1 and 3, together with the definition of $\mathfrak{b}$, immediately imply the following:

Corollary 1. *Let $F \subseteq K[X]$. Then there exists an exact 0-standard monomial cone decomposition Q of N_F. Moreover, if the polynomials in F are homogeneous, then $\deg(g) \leq \mathfrak{b}_{Q,0}$ for all g in the reduced Gröbner basis of F.*

The importance of exact cone decompositions stems from the fact that the Hilbert function of sets with an exact cone decomposition Q can be easily described in terms of the constants $\mathfrak{b}_{Q,i}$, for $1 \leq i \leq n+1$. This in turn enables us to obtain an upper bound for $\mathfrak{b}_{Q,0}$, which, by virtue of Corollary 1, is also an upper bound for the degrees of reduced Gröbner bases. Details follow in Section 4.

There is nothing special about the formal definitions of exact cone decompositions (constant `exact_decomp`), $\mathfrak{a}$ or $\mathfrak{b}$, hence we omit them here. Instead, we list some simple facts about $\mathfrak{b}$:

```
lemma b_decreasing: "i ≤ j ⟹ b qs j ≤ b qs i"

lemma b_zero: "qs ≠ [] ⟹ Max (poly_deg ∘ fst ∘ set qs) < b qs 0"

lemma b_card_X: "card X < i ⟹ b qs i = a qs"
```

The first lemma, $\mathfrak{b}$-*decreasing*, states that the sequence $(\mathfrak{b}_{Q,i})_{i\geq 0}$ is decreasing, $\mathfrak{b}$-*zero* states that $\mathfrak{b}_{Q,0}$ is an upper bound for the degrees of the tips h of pairs $(h,U) \in Q$, and $\mathfrak{b}$-*card-X* states that $\mathfrak{b}_{Q,i}$ stabilizes to $\mathfrak{a}_Q$ for sufficiently large i.

With respect to the proof of Theorem 3, the situation here parallels Section 3.2: there is an algorithm for transforming an arbitrary k-standard cone decomposition into an exact one, and this algorithm is implemented in function `exact` in the formalization. However, since the implementation of `exact` is more complicated than the one of `split` shown in Section 3.2, we omit it here. Using `exact`, the formal statement of Theorem 3 is distributed across several lemmas:

```
lemma exact:
  assumes "valid_decomp qs" "standard_decomp k qs"
  shows "valid_decomp (exact k qs)" "standard_decomp k (exact k qs)"
        "exact_decomp (exact k qs)"

lemma cone_decomp_exact:
  assumes "valid_decomp qs" "standard_decomp k qs" "cone_decomp T qs"
  shows "cone_decomp T (exact k qs)"

lemma Max_exact_ge:
  assumes "valid_decomp qs" and "standard_decomp k qs"
  shows "Max (poly_deg ∘ fst ∘ set qs) ≤ Max (poly_deg ∘ fst ∘ set (exact k qs))"
```

Note that `exact` not only takes a cone decomposition Q as input, but also a k for which Q is k-standard; if Q is not k-standard, nothing can be said about the cone decomposition returned by `exact`.

Finally, Corollary 1 is formalized as:

```
lemma normal_form_exact_decompE:
  assumes "F ⊆ P[X]"
  obtains qs where "valid_decomp qs" "standard_decomp 0 qs" "monomial_decomp qs"
    "cone_decomp (normal_form F ` P[X]) qs" "exact_decomp qs"
    "⋀g. (∀f∈F. homogeneous f) ⟹ g ∈ reduced_GB F ⟹ poly_deg g ≤ 𝔟 qs 0"
```

4 Obtaining the Degree Bound

We have almost all prerequisites to obtain the final degree bound. But still, we need one more definition:

Definition 5. *Let $b = (b_1, \ldots, b_{n+1})$ be a tuple of natural numbers, where as usual $n = |X|$. Then the* Hilbert polynomial *in b, denoted by $\overline{\varphi}_b(z)$, is defined as*

$$\overline{\varphi}_b(z) := \binom{z - b_{n+1} + n}{n} - 1 - \sum_{i=1}^{n} \binom{z - b_i + i - 1}{i}.$$

Since the upper entries of the binomial coefficients can be arbitrary numbers, the binomial coefficients are generalized *bin. coeffs.:* $\binom{a}{n} := \frac{a(a-1)(a-2)\ldots(a-(n-1))}{n!}$.

Abusing notation, we will abbreviate $\overline{\varphi}_{\mathfrak{b}_P}(z)$ simply by $\overline{\varphi}_P(z)$ if P is a cone decomposition.

Theorem 4 (see Section 7 in [4]). *Let P be an exact k-standard homogeneous cone decomposition of T. Then, if $z \geq \mathfrak{b}_{P,0}$, the Hilbert function of T equals the Hilbert polynomial of P, i. e. $\varphi_T(z) = \overline{\varphi}_P(z)$.*

The proof of Theorem 4 is based on the fact that the Hilbert function of T is just the sum of the Hilbert functions of the cones in a homogeneous cone decomposition P of T, and that this sum can be fully characterized by the $\mathfrak{b}_{P,i}$ if P is exact and k-standard. The rest is mere rewriting of (sums of) binomial coefficients by well-known binomial identities.

Now, let $F \subseteq K[X]$ be a finite set of homogeneous polynomials, and let $f \in F$ whose degree d is maximal among the degrees of the polynomials in F. Let Q be the exact 0-standard monomial cone decomposition of N_F whose existence is guaranteed by Corollary 1; since Q is 0-standard, we also know $\mathfrak{b}_{Q,n+1} = \mathfrak{a}_Q = 0$. Let $T \subseteq K[X]$ and P be an exact d-standard homogeneous cone decomposition of T such that $\langle F \rangle = \langle f \rangle \oplus T$, whose existence is guaranteed by Theorem 2 (strictly speaking, the cone decomposition from that theorem still has to be made exact using Theorem 3); similarly to $\mathfrak{b}_{Q,n+1}$, we know $\mathfrak{b}_{P,n+1} = d$. So, by Theorem 4 we obtain the following two identities for sufficiently large z:

$$\varphi_{\mathrm{N}_F}(z) = \overline{\varphi}_Q(z) = \binom{z+n}{n} - 1 - \sum_{i=1}^{n} \binom{z - \mathfrak{b}_{Q,i} + i - 1}{i}, \tag{2}$$

$$\begin{aligned} \varphi_{\langle F \rangle}(z) &= \overline{\varphi}_{\{(f,X)\}}(z) + \overline{\varphi}_P(z) \\ &= \binom{z-d+n-1}{n-1} + \binom{z+d-n}{n} - 1 - \sum_{i=1}^{n} \binom{z - \mathfrak{b}_{P,i} + i - 1}{i}. \end{aligned} \tag{3}$$

Together with (1) from Section 2.1 we therefore obtain the key identity

$$\begin{aligned} \binom{z+n-1}{n-1} = \ & \binom{z-d+n-1}{n-1} + \binom{z-d+n}{n} + \binom{z+n}{n} - 2 \\ & - \sum_{i=1}^{n} \left(\binom{z - \mathfrak{b}_{P,i} + i - 1}{i} + \binom{z - \mathfrak{b}_{Q,i} + i - 1}{i} \right) \end{aligned} \tag{4}$$

which holds for all sufficiently large z; but, since both sides of the equality are polynomials in z, and two polynomials agree everywhere if they agree on infinitely many arguments, we can conclude that the above identity actually holds for *all* z.

After a lengthy chain of simplifications and estimations of (4), which are in detail explained in [4], one can prove that $\mathfrak{b}_{P,j}$ and $\mathfrak{b}_{Q,j}$ satisfy $\mathfrak{b}_{P,j} + \mathfrak{b}_{Q,j} \leq \mathrm{Dube}_{n,d}(j)$ for $0 < j \leq n-1$, where $\mathrm{Dube}_{n,d}(j)$ is defined recursively as

$$\mathrm{Dube}_{n,d}(n-1) = 2d \tag{5}$$

$$\mathrm{Dube}_{n,d}(n-2) = d^2 + 2d \tag{6}$$

$$\mathrm{Dube}_{n,d}(j) = 2 + \binom{\mathrm{Dube}_{n,d}(j+1)}{2} + \sum_{i=j+3}^{n-1} \binom{\mathrm{Dube}_{n,d}(i)}{i-j+1}. \tag{7}$$

Note that $\mathrm{Dube}_{n,d}(j)$ is defined in terms of $\mathrm{Dube}_{n,d}(k)$ for *larger* k. In particular, $\mathfrak{b}_{P,1} + \mathfrak{b}_{Q,1} \leq \mathrm{Dube}_{n,d}(1) =: \mathrm{Dube}_{n,d}$, and since also $\mathfrak{b}_{Q,0} \leq \max\{\mathfrak{b}_{P,1}, \mathfrak{b}_{Q,1}\}$ can be proved, $\mathrm{Dube}_{n,d}$ is an upper bound for $\mathfrak{b}_{Q,0}$, too. Therefore, thanks to Corollary 1, $\mathrm{Dube}_{n,d}$ is the desired upper bound for the degrees of the polynomials in the reduced Gröbner basis of F.

Theorem 5 (Variant of Theorem 8.2. in [4]). *Let $F \subseteq K[X]$ be a finite set of homogeneous polynomials, and let d be the maximum degree of the polynomials in F. Then every g in the reduced Gröbner basis of F satisfies* $\deg(g) \leq \mathrm{Dube}_{n,d}$.

Remark 3. Some of the steps in the derivation above only hold if $d > 0$ and $n > 1$. The remaining cases of $d = 0$ and $n \leq 1$ can easily be handled separately, though: it is easy to see that d is a valid degree bound in these cases.

But even if the ideal under consideration is not homogeneous a result similar to Theorem 5 holds; in fact, it is even stronger, because it gives a bound on the *representation* of the Gröbner basis elements in terms of the polynomials in F:

Corollary 2 (Variant of Corollary 5.4. in [2]). *Let $F \subseteq K[X]$ be finite and let d be the maximum degree of the polynomials in F. Then there exists a Gröbner basis G of F such that every $g \in G$ can be written as $g = \sum_{f \in F} q_f f$ for some polynomials q_f, such that* $\deg(q_f f) \leq \mathrm{Dube}_{n+1,d}$ *for all $f \in F$. In particular, g also satisfies* $\deg(g) \leq \mathrm{Dube}_{n+1,d}$.

This corollary can be obtained easily from Theorem 5 by first *homogenizing* F, then computing the reduced Gröbner basis of the homogenized set, and eventually *dehomogenizing* this Gröbner basis. Only note that in the bound we get $n + 1$ instead of n, and that the bound holds for *some* Gröbner basis of F, not necessarily the reduced one.

Let us now turn to the formalization of the concepts and results presented above, starting with the Hilbert polynomial:

```
definition Hilbert_poly :: "(nat ⇒ nat) ⇒ int ⇒ int"
  where "Hilbert_poly b z = (let n = card X in
                               ((z - b (n + 1) + n) gchoose n) - 1 -
                                (∑i=1..n. (z - b i + i - 1) gchoose i))"
```

Note that `Hilbert_poly` is defined for, and returns, integers. Working with integers proved tremendously more convenient in the upcoming derivation than working with natural numbers. The infix operator `gchoose`, contained in the standard library of Isabelle/HOL, represents generalized binomial coefficients. Note that in a term like $z - b\ (n + 1) + n$, where z has type `int` and the other summands have type `nat`, the other summands are automatically coerced to type `int` by Isabelle/HOL.

So, in the formal development Theorem 4 corresponds to:

```
theorem Hilbert_fun_eq_Hilbert_poly:
  assumes "X ≠ {}" "valid_decomp ps" "hom_decomp ps" "cone_decomp T ps"
          "standard_decomp k ps" "exact_decomp ps" "𝔟 ps 0 ≤ z"
  shows "int (Hilbert_fun T z) = Hilbert_poly (𝔟 ps) z"
```

We spare the reader the formalization of the lengthy derivation of the degree bound, and only show the formal definition of $\mathrm{Dube}_{n,d}$ and the ultimate theorems:

```
function Dube_aux :: "nat ⇒ nat ⇒ nat ⇒ nat" where
  "Dube_aux n d j = (if j + 2 < n then
                       2 + ((Dube_aux n d (j + 1)) choose 2) +
                         (∑i=j+3..n-1. (Dube_aux n d i) choose (i - j + 1))
                     else if j + 2 = n then
                       d^2 + 2 * d
                     else 2 * d)"

definition Dube :: "nat ⇒ nat ⇒ nat"
  where "Dube n d = (if n ≤ 1 ∨ d = 0 then d else Dube_aux n d 1)"

theorem Dube:
  assumes "finite F" "F ⊆ P[X]" "∀f∈F. homogeneous f" "g ∈ reduced_GB F"
  shows "poly_deg g ≤ Dube (card X) (maxdeg F)"

corollary Dube_is_GB_cofactor_bound_explicit:
  assumes "finite F" "F ⊆ P[X]"
  obtains G where "is_Groebner_basis G" "ideal G = ideal F" "G ⊆ P[X]"
        "⋀g. g ∈ G ⟹ ∃q. g = (∑f∈F. q f * f) ∧
                     (∀f. poly_deg (q f * f) ≤ Dube (card X + 1) (maxdeg F))"
```

As can be seen, the statement of the formal Theorems *Dube* and *Dube-is-GB-cofactor-bound-explicit* correspond exactly to the statements of Theorem 5 and Corollary 2, respectively. Also, functions `Dube` and `Dube_aux` are effectively computable by means of Isabelle's code generator [5], meaning that for a concrete set F of polynomials one can compute an upper bound for the maximum degree of a Gröbner basis of F by a formally verified algorithm.

5 Conclusion

In the preceding sections we presented our Isabelle/HOL formalization of a degree bound for reduced Gröbner bases of homogeneous ideals, closely following [4]. In fact, the only substantial deviation from [4] is that there the constant $\mathrm{Dube}_{n,d}$ is further bounded from above by a nice closed form:

$$\mathrm{Dube}_{n,d} \leq 2\left(\frac{d^2}{2} + d\right)^{2^{n-2}}$$

for all $n \geq 2$. However, the proof of this inequality ([4, Lemma 8.1.]) contains two little mistakes: first, the recursive description of $\mathrm{Dube}_{n,d}(j)$ (cf. (7)) lacks the summand 2, and second, the author wrongly assumes that the sum $\sum_{i=j+3}^{n-1} \frac{2^{i-j}}{(i-j+1)!}$ is never greater than $1/2$; this is not true, e. g., for $n = 7$, $d = 2$ and $j = 1$, where that sum is $23/45$. To the best of our knowledge, these mistakes have remained unnoticed until now. Nevertheless, experiments indicate that the closed form *is* a valid upper bound, typically even *much* larger than the value of

$\text{Dube}_{n,d}$ for concrete n and d,[6] but a rigorous proof of this claim must be left as future work. For our purpose the recursively defined function $\text{Dube}_{n,d}$ is absolutely sufficient anyway: it is (easily) computable and, as said before, typically gives much better bounds.

Formalizing the results presented in this paper was not entirely trivial, despite the fact that Dubé's original paper is written very well and explains every step of the proof in detail. The reason is that during the formalization process we had to backtrack a design choice we made at the beginning: at first, we wanted to simplify matters and therefore did not base the development on polynomials and their leading power-products, but on power-products directly; this is reflected, for instance, in function `split`, which only takes power-products as input and originally returned power-products rather than polynomials, too. Later it turned out, however, that some theorems simply cannot be proved without any reference to polynomials and ideals thereof, not even after adjusting their statements to fit into the 'power-products-only' framework; Theorem 2 serves as a good example. Fortunately the wrong design choice could be corrected with only moderate effort and did not cost too much time in the end. This also owes to the large arsenal of sophisticated proof methods offered by the underlying Isabelle system, which enabled us to construct rather abstract proofs that remained valid even after replacing power-products by polynomials (think of `auto`, for instance). In general, it was not necessary to develop any new proof methods or other tools for completing the formalization.

The total number of lines of proof of the formalization is more than 11000, which is quite significant. Note that [8] also contains other material this paper is not concerned with and which is therefore not counted in the lines of proof; in particular, it contains a formalization of the Macaulay-matrix-based approach to Gröbner bases mentioned in the introduction (theory 'Groebner_Macaulay'), described in [7]. Formalizing the theory took roughly 270 working hours.

Having settled the general case of arbitrary input sets F, one could now try to look at special cases that admit tighter degree bounds. For instance, Wiesinger-Widi in [14] restricts herself to sets F consisting only of two binomials and derives significantly better bounds there. Formalizing her results is a challenging task, though, especially since her proofs make use of fairly different techniques than Dubé's proof presented here.

Acknowledgments. I thank the anonymous referees for their valuable comments.

References

1. Adams, W.W., Loustaunau, P.: An Introduction to Gröbner Bases, Graduate Studies in Mathematics, vol. 3. AMS (1994)
2. Aschenbrenner, M., Leykin, A.: Degree Bounds for Gröbner Bases in Algebras of Solvable Type. Journal of Pure and Applied Algebra **213**, 1578–1605 (2009). https://doi.org/10.1016/j.jpaa.2008.11.022

[6] Although asymptotically $\text{Dube}_{n,d}$ also has to grow double exponentially in n, as shown in [10].

3. Buchberger, B.: Ein Algorithmus zum Auffinden der Basiselemente des Restklassenrings nach einem nulldimensionalen Polynomideal. Ph.D. thesis, Mathematisches Institut, Universität Innsbruck, Austria (1965), English translation in Journal of Symbolic Computation 41(3–4), 475–511 (2006)
4. Dubé, T.W.: The Structure of Polynomial Ideals and Gröbner Bases. SIAM Journal on Computing **19**(4), 750–773 (1990). https://doi.org/10.1137/0219053
5. Haftmann, F.: Code Generation from Isabelle/HOL Theories, `http://isabelle.in.tum.de/dist/Isabelle2018/doc/codegen.pdf`, part of the Isabelle documentation
6. Immler, F., Maletzky, A.: Gröbner Bases Theory. Archive of Formal Proofs (2016), `http://isa-afp.org/entries/Groebner_Bases.html`, formal proof development
7. Maletzky, A.: Gröbner Bases and Macaulay Matrices in Isabelle/HOL. Tech. rep., RISC, Johannes Kepler University Linz, Austria (2018), `http://www.risc.jku.at/publications/download/risc_5814/Paper.pdf`
8. Maletzky, A.: Isabelle/HOL Formalization of Advanced Gröbner Bases Material (2019), `https://github.com/amaletzk/Isabelle-Groebner/`
9. Maletzky, A., Immler, F.: Gröbner Bases of Modules and Faugère's F_4 Algorithm in Isabelle/HOL. In: Rabe, F., Farmer, W., Passmore, G., Youssef, A. (eds.) Intelligent Computer Mathematics (Proceedings of CICM 2018, Hagenberg, Austria, August 13-17). Lecture Notes in Computer Science, vol. 11006, pp. 178–193. Springer (2018). https://doi.org/10.1007/978-3-319-96812-4_16
10. Mayr, E.W., Meyer, A.R.: The Complexity of the Word Problems for Commutative Semigroups and Polynomial Ideals. Advances in Mathematics **46**(3), 305–329 (1982). https://doi.org/10.1016/0001-8708(82)90048-2
11. Nipkow, T., Paulson, L.C., Wenzel, M.: Isabelle/HOL—A Proof Assistant for Higher-Order Logic, LNCS, vol. 2283. Springer (2002). https://doi.org/10.1007/3-540-45949-9
12. Sternagel, C., Thiemann, R., Maletzky, A., Immler, F.: Executable Multivariate Polynomials. Archive of Formal Proofs (2010), `http://isa-afp.org/entries/Polynomials.html`, formal proof development
13. Wenzel, M.: The Isabelle/Isar Reference Manual (2018), `https://isabelle.in.tum.de/dist/Isabelle2018/doc/isar-ref.pdf`, part of the Isabelle documentation
14. Wiesinger-Widi, M.: Gröbner Bases and Generalized Sylvester Matrices. Ph.D. thesis, RISC, Johannes Kepler University Linz, Austria (2015), `http://epub.jku.at/obvulihs/content/titleinfo/776913`