Learning Selection Strategies in Buchberger's Algorithm: Supplementary Material

June 30, 2020

1. Experimental methods

The code used to generate all statistics and results for this paper is available at

https://github.com/dylanpeifer/deepgroebner

with selected computed statistics and training run data at

https://doi.org/10.5281/zenodo.3676044

Statistics for First, Degree, Normal, Sugar, and Random were generated using Macaulay2 1.14 on Ubuntu 18.04 while agent training was performed on c5n.xlarge instances from Amazon Web Services using the Ubuntu 18.04 Deep Learning AMI.

There were four primary training settings corresponding to the four distributions in Table 3 from Section 5. In each setting we performed three complete runs using the parameters from Table 1. Model weights were saved every 100 epochs, and a single model was selected from the available runs and save points in each setting based on best smoothed training performance.

Table 1. Hyperparameters for primary evaluation runs.

HYPERPARAMETER	VALUE
γ FOR GAE	0.99
λ FOR GAE	0.97
ϵ FOR PPO	0.2
OPTIMIZER	ADAM
LEARNING RATE	0.0001
MAX POLICY UPDATES PER EPOCH	80
POLICY KL-DIVERGENCE LIMIT	0.01
HIDDEN LAYERS	[128]
VALUE FUNCTION	DEGREE AGENT
EPOCHS	2500
EPISODES PER EPOCH	100
MAX EPISODE LENGTH	500

Trained models were then evaluated on new sets of ideals to produce Table 3 in Section 5, Table 5 in Section 5.3, and Figure 6 in Section 5.3. The three models from Table 4 in Section 5.2 were selected and evaluated in the same way, but were trained with their respective modifications.

2. Hyperparameter tuning

In addition to the main evaluation runs for this paper, we performed a brief hyperparameter search in two stages. Both stages were trained in the 3-20-10 weighted distribution as it gives the fastest training runs.

In the first stage we varied the parameters γ in $\{1.0, 0.99\}$, λ in $\{1.0, 0.97\}$, learning rate in $\{10^{-3}, 10^{-4}, 10^{-5}\}$, and network as a single hidden layer of 128 units or two hidden layers of 64 units. Three runs were performed on each set of parameters, for a total of 72 runs. The pairs left value function was used in this search instead of the degree agent, as it leads to significantly faster training. Learning rates of 10^{-4} showed best performance, though rates of 10^{-5} were still improving at the end of the runs. Changes in γ, λ, and network did not consistently change performance.

In the second stage we varied just the network shape. Single hidden layer networks were tried with 4, 8, ..., 256 units and two hidden layer networks were tried with 4, 8, ..., 256 hidden units in each layer. One run was performed on each network, for a total of 14 runs. Results showed significant improvement in using at least 32 hidden units and no major differences between one and two hidden layers. Small models were also surprisingly effective, with the model with a single hidden layer of 4 units achieving mean performance of around 100 polynomial additions during training, compared to Degree selection at 136 and our full model at 85.6.

3. Complexity analysis of Buchberger's algorithm

In this section, we consider upper and lower bounds for the complexity of computing a Gröbner basis of an ideal in a polynomial ring, and also describe what happens in the "generic" (random) case, giving more detail than in the paper proper. We first consider the maximum degree of a Gröbner basis element, then describe how that gives bounds for the size of a minimal reduced Gröbner basis.

Let $I = \langle f_1, \dots, f_s \rangle \subset S = k[x_1, \dots, x_n]$, be an ideal, with each polynomial f_i of degree $\leq d$. If $>$ is

a monomial order, and $G = GB_>(I) = \{g_1, \ldots, g_r\}$ is a Gröbner basis of I, G is called a *minimal and reduced* Gröbner basis if the lead coefficient of each g_i is one, and no monomial of g_i is divisible by $LT(g_j)$, for $i \neq j$. Given any Gröbner basis, it is easy to modify G to obtain a minimal and reduced Gröbner basis of I. Given the monomial order, each ideal I has precisely one minimal reduced Gröbner basis with respect to this order. We define $\deg_{\max}(GB_>(I)) := \max(\deg g_1, \ldots, \deg g_r)$, where $G = GB_>(I) = \{g_1, \ldots, g_r\}$ is the unique minimal reduced Gröbner basis of I for the order $>$.

Upper bounds

We have the following upper bound for $\deg_{\max}(GB_>(I))$.

Theorem 1 ((Dubé, 1990)). *Given I as above, then*

$$\deg_{\max}(GB_>(I)) \leq 2\,(\frac{d^2}{2} + d)^{2^{n-1}}$$

If I is homogeneous (i.e. each polynomial f_i is homogeneous), we may replace the $n-1$ by $n-2$ in this bound.

Such a bound is called double exponential (in the number of variables). This result seems to give an incredibly bad bound, but it is unfortunately fairly tight, which we will discuss next.

Lower bounds

All known double exponential examples (e.g. (Bayer & Stillman, 1988), (Koh, 1998), (Mora, 2005)) are based essentially on the seminal and important construction of (Mayr & Meyer, 1982). Each is a sequence of ideals J_n where the n-th ideal J_n is in roughly $10n$ or $20n$ variables, generated in degrees $\leq d$, where each f_i is a *pure binomial* (i.e. a difference of two monomials). The following version of (Koh, 1998) is a sequence of ideals generated by *quadratic* pure binomials.

Theorem 2 ((Koh, 1998)). *For each $n \geq 1$, there exists an ideal J_n, generated by quadratic homogeneous pure binomials in $22n-1$ variables such that for any monomial order $>$,*

$$2^{2^{n-1}-1} \leq \deg_{\max}(GB_>(J_n))$$

(Koh, 1998) shows that there is a minimal syzygy in degree $2^{2^{n-1}}$. It is well known (see e.g. (Mora, 2005), section 38.1) that this implies that there must be a minimal Gröbner basis element of degree at least half that, giving the stated bound. Thus there are examples of ideals whose Gröbner basis has maximum degree bounded below by roughly $2^{2^{n/22}}$ (where now n is the number of variables). Some of the other modifications of the Mayr-Meyer construction have slightly higher lower bounds (e.g.$2^{2^{n/10}}$).

Better bounds

Given these very large lower bounds, one might conclude that Gröbner bases cannot be used in practice. However, in many cases, there exist much lower upper bounds for the size of a grevlex Gröbner basis. The key is to relate these degree bounds to the *regularity* of the ideal.

Given a homogeneous ideal $I = \langle f_1, \ldots, f_s \rangle \subset S = k[x_1, \ldots, x_n]$, with each polynomial of degree $\leq d$, several notions which often appear in complexity bounds and are also useful in algebraic geometry are:

- the dimension $dim(I)$ of I.
- the depth $depth(I)$. This is an integer in the range $0 \leq depth(I) \leq dim(I)$. In many commutative algebra texts, this is denoted as $depth(S/I)$, not $depth(I)$, but in (Mora, 2005), $depth(I)$ is the notation. This is the length of a maximal S/I-regular sequence in $(x_1, \ldots, x_n)$.
- the (Castelnuovo-Mumford) regularity, $reg(I)$ of the ideal I, see (Eisenbud, 1995) or (Mora, 2005).

The regularity $reg(I)$ should be considered as a measure of complexity of the ideal.

Generic change of coordinates

Let's consider a homogeneous, linear, change of coordinates $\phi = \phi_A$, where $A \in k^{n \times n}$ is a square n by n matrix over k, with

$$\phi_A(x_i) = \sum_{j=1}^{n} A_{ij} x_j.$$

Let $\phi_A(I) := \{f(\phi_A(x_1), \ldots, \phi_A(x_n)) \mid f \in I\}$ be the ideal under a change of coordinates. Consider the n^2-dimensional parameter space V (where a point $A = (A_{ij})$ of V corresponds to a homogeneous linear change of coordinates ϕ_A). It turns out that there is a polynomial F in the polynomial ring (with n^2 variables) $k[A_{ij}]$, such that for all points $A \in V$ such that $F(A) \neq 0$, then $LT_{grevlex}(\phi_A(I))$ is the same ideal. This monomial ideal is called the *generic initial ideal* of I (in grevlex coordinates), and is denoted by $gin(I)$. Basically, for a random homogeneous linear change of coordinates, one always gets the same size Gröbner basis, with the same lead monomials.

Define $G(I)$ to be the maximum degree of a minimal generator of $gin(I)$. This is the maximum degree of an element of the unique minimal and reduced Gröbner basis of the ideal $\phi_A(I)$ under almost all change of coordinates ϕ_A (i.e. those for which $F(A) \neq 0$).

The reason this is important is that we have more control over Gröbner bases in generic coordinates. For instance

Theorem 3 ((Bayer & Stillman, 1987)). *If the base field is infinite, then*

$$reg(I) = reg(gin(I))$$

If the characteristic of k is zero, then

$$G(I) = reg(I)$$

If the characteristic of k is positive, then

$$\frac{1}{n} reg(I) \leq G(I) \leq reg(I)$$

It is know that $reg(I) \leq reg(LT_>(I))$, for every monomial order $>$. This result states that in fact in generic coordinates, equality is obtained for the grevlex order. The Gröbner basis, after a random change of coordinates, always has maximum degree at most $reg(I)$.

In particular, if an ideal I has small regularity, as often happens for ideals coming from algebraic geometric problems, then the corresponding Gröbner basis in grevlex order will have much smaller size than the double exponential upper bounds suggest.

Theorem 4. *If the homogeneous ideal $I = \langle f_1, \ldots, f_s \rangle \subset S$ has $\dim(I) = depth(I)$ (this includes the case when $dim(I) = 0$, then*

$$reg(I) \leq (d-1)\min(s, n - \dim(I))) + 1$$

This follows from two basic facts about regularity: First, if $\dim I = 0$, then the regularity of I is the first degree m such that the degree m polynomials in I consist of all degree m polynomials. Second, if I has depth r and $y_1, \ldots, y_r$ is a regular sequence of linear forms mod I, then the regularity of I is the regularity of the ideal $\overline{I} := IS/(y_1, \ldots, y_r)$. Since the depth and dimension of I are equal, the ideal $\overline{I}$ is of dimension 0, and contains a complete intersection of polynomials each of degree d. This implies by a Hilbert function argument, or by the Koszul complex, that the regularity of $\overline{I}$ is at most $(d-1)(n-r)+1$ (see (Eisenbud, 1995) for these kinds of arguments).

This implies that $G(I) \leq (d-1)\min(s, n - \dim(I)) + 1 \leq dn$, a dramatic improvement on the double exponential bounds!

Ideals generated by random, or generic, polynomials

What happens for random homogeneous ideals generated by s polynomials each of degree d? For fixed n, d, s, the space of possible inputs, i.e., the space V of coefficients for each of the s generators, is finite dimensional. There is a subset $X \subset V$, a closed algebraic set (so having measure zero, if the base field is $\mathbb{R}$ or $\mathbb{C}$), such that for any point outside X, the corresponding ideal I satisfies $\dim(I) = depth(I)$, and therefore,

$$G(I) \leq (d-1)\min(s, n) + 1.$$

In characteristic zero, equality holds.

If instead of homogeneous ideals, we consider random inhomogeneous ideals, generated by s polynomials each of degree d. The same method holds: the homogenization of these polynomials puts us into the situation in the previous paragraph. Therefore for such inhomogeneous ideals, whose coefficient point is outside of X, then the ideal J generated by the homogenization of the f_i with respect to a new variable satisfies $\dim(J) = depth(J)$, and therefore,

$$G(I) = G(J) \leq (d-1)\min(s, n+1) + 1.$$

In characteristic zero, equality holds.

Bounds on the size of the reduced minimal Gröbner basis

In the unique reduced minimla Gröbner basis of an ideal I, there can not be two generators with identical lead monomials. It follows that if all generators in this Gröbner basis have degree $\leq D$, then there are at most

$$\begin{aligned} \#\{\text{monomials of degree } \leq D\} &= \begin{pmatrix} D+n \\ n \end{pmatrix} \\ &= \mathcal{O}\left((n+D)^{\min(n,D)}\right) \end{aligned}$$

generators in the Gröbner basis. If one combines this with the upper bound above on the maximum degree, $D = (d-1)\min(s, n+1) + 1$, one finds the following upper bound on the size of the minimal reduced Gröbner basis of a generic ideal generated by s polynomials of degree $\leq d$ in n variables:

$$\#\{\text{Gröbner basis generators}\} \leq \mathcal{O}\left((n+1)^n d^n\right),$$

where the simplification comes from approximating $\min(s, n+1) \leq n+1$, so that our bound is independent of s, and assuming $d \geq 2$.

References

Bayer, D. and Stillman, M. A criterion for detecting m-regularity. *Invent. Math.*, 87(1):1–11, 1987.

Bayer, D. and Stillman, M. On the complexity of computing syzygies. *J. Symbolic Comput.*, 6(2-3):135–147, 1988.

Dubé, T. W. The structure of polynomial ideals and Gröbner bases. *SIAM J. Comput.*, 19(4):750–775, 1990.

Eisenbud, D. *Commutative algebra*, volume 150 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995. With a view toward algebraic geometry.

Koh, J. Ideals generated by quadrics exhibiting double exponential degrees. *J. Algebra*, 200(1):225–245, 1998.

Mayr, E. W. and Meyer, A. R. The complexity of the word problems for commutative semigroups and polynomial ideals. *Adv. in Math.*, 46(3):305–329, 1982.

Mora, T. *Solving polynomial equation systems. II*, volume 99 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2005.