On the complexity of Gröbner basis computation of semi-regular overdetermined algebraic equations
Magali Bardet, Jean-Charles Faugère, Bruno Salvy

Abstract

We extend the notion of regular sequence ([Mac16]) to overdetermined system of algebraic equations. We study generic properties of Gröbner bases and analyse precisely the behavior of the $F_5$ [Fau02] algorithm. Sharp asymptotic estimates of the degree of regularity are given.

We consider polynomials $(f_1, \ldots, f_m)$ in $k[x_1, \ldots, x_n]$ where $k$ is a field. In this extended abstract, we restrict attention to homogeneous polynomials. We denote by $d_i$ the total degree of $f_i$.

Introduction

Gröbner bases [Buc65, CLO98] are a fundamental tool to study algebraic equations in theory and practice. Complexity of Gröbner bases has been the object of extensive studies. Since Gröbner bases can be used to solve polynomial systems, their complexity is at least that of polynomial system solving. It turns out that it is not difficult to encode NP-complete problems (Knapsack problem, $k$-SAT, ...) into polynomial systems; hence polynomial system solving is hard which shows that the worst-case complexity cannot be expected to be good.

Actually, while the worst-case is at least “double exponential”\footnote{more precisely $\text{cste}2^{2^\frac{d}{2}}$ where $n$ is the number of variables}, the generic behaviour is much better. For instance, if the algebraic system has only a finite number of common zeros at infinity, then, its Gröbner Basis for any ordering may be computed in a time polynomial in $d^n$ where $d = \max_i d_i$. In that case, for the degree–reverse–lexicographical (DRL) ordering, the highest degree of elements of the Gröbner basis is bounded very precisely [Laz83, Giu94] by

$$\text{The Macaulay bound : } \sum_{i=1}^n (d_i - 1) + 1.$$  \hspace{1cm} (1)

These bounds should be compared with Bézout's theorem, stating that the number of solutions, when finite, is bounded by $\Pi_i d_i$, and is exactly $\Pi_i d_i$ in the homogeneous case. This picture leads to natural questions that are (partially) adressed in the full version of the article:

Where are “random” systems? What is their complexity? What about overdetermined systems?

The goal of the article is to extend the bound (1) when the number of equations is larger than the number of variables and to derive sharp bounds on the complexity for the $F_5$ algorithm. The interest of overdetermined systems is not purely academic: many systems appearing in cryptography have been based on the problem of solving a system of algebraic equations over the finite field $\mathbb{F}_2$, and in many cases the interesting solutions are only solutions in $\mathbb{F}_2$ and not in its algebraic closure: one has to solve the original system of, say $m$, equations over $\mathbb{F}_2$ together with the field equations $x_i(x_i - 1) = 0$ ($i = 1, \ldots, n$). Thus the total number of equations is $m + n$. Other applications are: error correcting codes (decoding of cyclic codes), robotic, calibration, ...

Regular systems

The $F_5$ algorithm was designed so that it ensures no “useless” reduction to 0 when the input system is regular. We recall the definition of regularity (regular sequence, Macaulay):

\footnote{more precisely $\text{cste}2^{2^\frac{d}{2}}$ where $n$ is the number of variables}
Definition 1. \((f_1, \ldots, f_m)\) is regular if for all \(i = 1, \ldots, m\), \(f_i\) is not a zero-divisor in the quotient ring \(k[x_1, \ldots, x_n]/(f_1, \ldots, f_{i-1})\). In other words if there exists \(g\) such that
\[
gf_i \in \text{Ideal}(f_1, \ldots, f_{i-1})
\]
then \(g\) is also in \(\text{Ideal}(f_1, \ldots, f_{i-1})\).

Classical properties of regular systems are:

Theorem 2. (i) \((f_1, \ldots, f_m)\) is regular if and only if its Hilbert series is given by
\[
H(t) = \frac{\prod_{j=1}^{n} (1 - t^{d_j})}{(1 - t)^n}
\]
(ii) after a generic linear change of variables, the highest degree of elements of a Gröbner basis for the DRL order is less than
\[
\sum_{i=1}^{n} (d_i - 1) + 1
\]

Semi-Regular systems

Unfortunately regular systems do not exist when the number of polynomials is larger than the number of variables. We have to modify slightly the definition of regularity:

Definition 3. A zero-dimensional overdetermined system \((f_1, \ldots, f_m)\) \((m \geq n)\) is \(d\)-regular when for all \(i = 1, \ldots, m\), if there exists \(g\) such that
\[
\deg(g) < d - d_i \quad \text{and} \quad gf_i \in \text{Ideal}(f_1, \ldots, f_{i-1})
\]
then \(g\) is also in \(\text{Ideal}(f_1, \ldots, f_{i-1})\).

For instance, a quadratic system of equations is 2-regular if the equations are linearly independent. The maximum expected value of \(d\) is given by the following definition:

Definition 4. We define the degree of regularity of a zero dimensional ideal \(I = \text{Ideal}(f_1, \ldots, f_m)\) \((m \geq n)\) by
\[
d_{\text{reg}} = \min \left\{ d \geq 0 \mid \dim_k(\{f \in I, \quad \deg(f) = d\}) = \binom{n + d - 1}{d} \right\}
\]
This definition implies that for any monomial ordering refining the degree, all monomials in degree \(d_{\text{reg}}\) are leading terms for an element of the ideal. Thus \(d_{\text{reg}}\) is clearly an upper bound on the degree of the elements of a Gröbner basis for such a monomial ordering.

Definition 5. A \(d_{\text{reg}}\)-regular system is called semi-regular.

Thus when \(m = n\) a regular (zero-dimensional) system is also semi-regular. The following proposition gives a way to compute \(d_{\text{reg}}\) efficiently:

Proposition 6. For a semi-regular system with \(m \geq n\) polynomials, the degree of regularity is the index of the first nonpositive \((\leq 0)\) coefficient in the series \(H(t)\).

We can now state one of the main results of this article:

Theorem 7. For a \(d\)-regular system, there is no reduction to 0 in the algorithm \(F_5\) for degrees smaller than \(d\). Moreover, for a semi-regular system, the total number of arithmetic operations in \(k\) performed by \(F_5\) is bounded by
\[
O \left( \binom{n + d_{\text{reg}}}{n} \omega \right)
\]
Where the exponent \(\omega < 2.39\) is the exponent in the complexity of matrix multiplication.
Asymptotic Analysis

The method is the following: the $k$-th coefficient of the series $H(t)$ is given by the Cauchy integral representation

$$I(k) = \frac{1}{2\pi i} \oint \frac{\prod_{i=1}^{m} (1 - t^{d_i})}{(1 - t)^n} \frac{dt}{tk+1} \quad (3)$$

A preliminary analysis reveals that the degree of regularity grows roughly linearly with $n$, that is to say $\lambda = \frac{d_{\text{reg}}}{n}$ is equivalent to some constant at infinity. The analysis is then based on computing the asymptotic expansion of $I(\lambda n)$ for fixed $\lambda$, and then determining an asymptotic expansion $\lambda(n)$ that makes this behaviour vanish asymptotically.

By using the saddle-point method, we are able to prove:

**Theorem 8.** The degree of regularity of a semi-regular system of $m = n + k$ homogeneous polynomials of degree $d_1, \ldots, d_{n+k}$ in $n$ variables behaves asymptotically like

$$d_{\text{reg}} = \sum_{i=1}^{m} d_i - 1 + \frac{1}{2} - \alpha_k \sqrt{\sum_{i=1}^{m} \frac{(d_i^2 - 1)}{6} + O(1)} \text{ when } n \to \infty$$

where $\alpha_k$ is the largest zero of the $k$-th Hermite polynomial.

For instance, for quadratic systems we have $d_{\text{reg}} \approx \frac{m - \alpha_1 \sqrt{2m}}{2}$. When $m = n + 1$, $\alpha_1 = 0$ and the result found is in agreement with the exact result due to Szanto [Sza04].

A similar analysis can be done when $m = \alpha n$ ($\alpha \geq 1$ being fixed); using the coalescent saddle points method a full asymptotic expansion can be computed:

**Theorem 9.** The degree of regularity of a semi-regular system of $m = \alpha n$ homogeneous polynomials of degree $d_1, \ldots, d_{\alpha n}$ in $n$ variables behaves asymptotically like

$$d_{\text{reg}} = \phi(\rho)n - a_1 \sqrt{\left(-\frac{1}{2}\phi''(\rho)\rho^2\right)} n + \cdots \text{ when } n \to \infty$$

where $\phi(z) = \frac{1}{z^2} - \frac{1}{n} \sum_{i=1}^{m} \frac{d_i^{\alpha_i}}{1 - z^{d_i}}$, $\rho$ is the zero of $\phi'(z)$ that minimize $\phi(\rho) > 0$ (an algebraic number) and $a_1$ is the largest zero of the classical Airy function.

For instance for quadratic equations and $m = 2n$ we can greatly improve the Macaulay bound $d_{\text{reg}} \leq n + 1$ with the new estimate:

$$d_{\text{reg}} = 0.0858 n + 1.04 \frac{n^4}{n^3} - 1.47 + \frac{1.71}{n^2} + O\left(\frac{1}{n^3}\right)$$

Extensions

The full version of the article includes several extensions. We give a definition of semi-regular systems for nonhomogeneous polynomials and we can deduce from our analysis a bound on the complexity of the Gröbner basis computation. Another extension is the boolean case: application of Gröbner bases in cryptography involves overdetermined systems over the field $\mathbb{F}_2$ and moreover the solutions themselves are sought in $\mathbb{F}_2$. In that case, it is convenient to modify the algorithm $F_5$ so that extra “useless” lines coming from the new syzygy $f^2_i = f_i$ are not computed. This results in an efficient algorithm that has been used to break a cryptographic challenge [FJ03]. The analysis proceeds as before, the degree of regularity being now the first nonpositive coefficient in the series $\prod_{i=1}^{n} \frac{1}{(1+\epsilon x_i)}$. A complexity bound for solving algebraic system using the algorithm XL can be derived from this analysis and the link between XL and Gröbner bases [AFI+04].

References:
References


