BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 38, Number 4, Pages 481–488
S 0273-0979(01)00915-6
Article electronically published on June 12, 2001

Gröbner deformations of hypergeometric differential equations, by M. Saito, B. Sturmfels, and N. Takayama, Springer, New York, 2000, viii+254 pp., \$42.00, ISBN 3-540-66065-8

This book is an introduction to new computational methods in the theory of linear PDE. To explain the terminology, the "old" methods, well predating the advent of computers, are concerned with approximate *numerical* solutions based on difference approximations to differential equations. The underlying mathematical point of view here is that a function is given by a table of its values. In contrast, the new computational methods forming the subject of this book are *symbolic*, i.e., based on the idea that a function is best given by a formula, e.g., as an explicit polynomial or power series. The goal then is to obtain an algorithm for finding the formula for the solution (e.g., finding the coefficients of its power series expansion) rather than for determining its values at a series of points.

The paradigm of symbolic, as opposed to numerical, computation first demonstrated its power in algebra, where it allowed a computer to handle the "abstract" algebra of rings and modules while the numerical approach was restricted to the more "classical" algebra of numbers and equations. This development was of fundamental importance for algebraic geometry since sophisticated questions about algebraic varieties were opened up for direct computer-assisted testing. The key mathematical concept here is that of Gröbner bases (see below). There are several excellent books on this subject, such as [1], [11].

The book under review is the first monograph in English on the application of Gröbner bases to differential equations and should be welcomed most enthusiastically. The authors give both a general treatment of holonomic systems of PDE and Gröbner bases methods for their solutions and illustrate the computational methods on a particular class of such systems coming from the theory of hypergeometric functions. In what follows we discuss the main concepts involved in more detail.

1. Holonomic systems and D-modules

As is well known, the theory of partial differential equations has a quite different flavor from that of ordinary differential equations. Typically, solutions form infinite-dimensional spaces, and also the very concept of a solution is open to interpretation (e.g., we can look for analytic, C^∞, distribution solutions, etc., and the corresponding solution spaces are often different). It was known on some level for a long time that the classical flavor of the theory can be regained by considering not one differential equation for a function of n variables (as is typically done in PDE) but systems of several (typically n) such equations. To give an elementary example,

$$u_y - u_{xx} = -10u, \quad u_{xx} - u_{yy} = u \tag{1}$$

is a system of two equations (one parabolic, one hyperbolic) on a function $u(x, y)$. While each of them has a complicated space of solutions, the space of simultaneous

2000 *Mathematics Subject Classification.* Primary 13P10, 14Qxx, 16S32, 33Cxx, 34Exx, 35Axx, 68W30.

solutions (understood in any reasonable sense) is 4-dimensional and spanned by the analytic functions $\exp(\xi x + \eta y)$ where (ξ, η) is one of the 4 points of intersection of the parabola $\eta = \xi^2 - 10$ and the hyperbola $\xi^2 - \eta^2 = 1$. This is a most direct instance of the method of characteristics which is equally important for ODE and PDE.

Systems which have "just the right number" of equations to ensure that the space of solutions is finite-dimensional are now called holonomic. While their theory goes back at least to E. Cartan, the special study of the linear case was begun by D. Quillen and D.C. Spencer in the 60's and then cast into the framework of D-modules by I.N. Bernstein and M. Kashiwara. To explain the essence of the D-module approach, consider a general system of linear PDE on a function $u(x_1, .., x_n)$ of n variables:

$$P_i\, u = 0, \quad i = 1, ..., r. \tag{2}$$

Here P_i are linear differential operators whose coefficients we, for simplicity, assume to be polynomials. All possible such operators form a noncommutative ring D (the Weyl algebra) with respect to composition. It is clear that any solution of (2) also satisfies $Ru = 0$ where R is of the form $\sum Q_i P_i$ with $Q_i \in D$ being arbitrary operators. Such operators form a left ideal $I = \sum DP_i$ in D, and one can associate to (2) the left D-module $M = D/I$. There are several advantages in doing this rather than keeping the system (2) as the basic object:

- A solution of (2) is the same as a D-module homomorphism $M \to \mathcal{O}$ where $\mathcal{O}$ is the space of functions we are considering (it is a left D-module as differential operators act on it).
- One can use the module M (or the ideal I) to decide whether (2) is holonomic and to find the dimension (called the holonomic rank) of the space of solutions near a generic point. This is again done by the method of characteristics. For the case of linear PDE with variable coefficients this method consists in associating to a differential operator R its *highest symbol* $S(R)$ which is a polynomial in x_i, ξ_i, $i = 1, ..., n$, obtained from R by replacing each $\partial/\partial x_i$ by ξ_i and then retaining only the terms of the highest possible total degree in the ξ_i. This degree is just the order $\mathrm{ord}(P)$ of P in the usual sense. We have then the two fundamental facts:

(A) The system (2) is holonomic if and only if its *characteristic variety* $\mathcal{C} \subset \mathbf{C}^{2n}$, defined as the set of common zeroes of $S(R), R \in I$, has dimension n. The space of solutions near a generic point is nonzero if and only if $\mathcal{C}$ contains the n-dimensional subvariety $\mathbf{C}^n$ given by $\xi_i = 0$, $i = 1, ...n$.

(B) As $\mathcal{C}$ is given by a system of algebraic equations, its component $\mathbf{C}^n$ comes with a natural multiplicity μ. This μ is equal to the holonomic rank.

Note that none of this will be true in general if we attempt to define $\mathcal{C}$ by the highest symbols of the original equations P_i alone. Indeed, it can happen that

$$\mathrm{ord}\left(\sum Q_i P_i\right) < \max_i(\mathrm{ord}(Q_i P_i)) \tag{3}$$

and thus $S\left(\sum Q_i P_i\right)$ need not lie in the ideal generated by the $S(P_i)$. For example, it is perfectly possible that a system of $r < n$ equations (2) on a function of n variables is already holonomic or even contradictory, but we would not be able to see this from the $S(P_i)$.

2. Gröbner bases

By twisting a little the historical order, one can present the concept of Gröbner bases (introduced by B. Buchberger) as a natural development of the above analysis of systems of PDE. More precisely, a set $\{P_1, ..., P_r\}$ of generators of a left ideal $I \in D$ is called a Gröbner basis if the highest symbols $S(P_i)$ generate the ideal formed by all the $S(R), R \in I$. In other words, we just try to avoid unpleasant consequences of phenomena such as (3). For the purpose of designing actual algorithms, however, it is important to generalize the concept of order and of the highest symbol. This can be done in two ways:

Version 1: Associate to every variable x_i and to every derivative $\partial_i = \partial/\partial x_i$ some weights and calculate the order of the operator according to these weights. (The usual concept of order corresponds to ∂_i having weight 1 and x_i weight 0.)

Version 2: Choose a total ordering (enumeration) of differential monomials $x^I \partial^J$, starting from $1 = x^0 \partial^0$, in a way compatible with the product. Such is, for example, an appropriate lexicographic ordering, while a system of weights gives usually only a partial order. Then define $\mathrm{ord}(P)$ to be the maximum of the numbers of its differential monomials with respect to the chosen enumeration.

In both cases ord gives a filtration of the ring D, and the highest symbol is defined in an obvious way as an element of the associated graded ring. Gröbner bases are then defined similarly to the above. Version 2 is important for computations since the symbol is always a single differential monomial (with a coefficient).

Historically, this approach was first applied to the polynomial algebra $\mathbf{C}[x_1,...,x_n]$ instead of D, and Buchberger devised an algorithm to construct a Gröbner basis for a polynomial ideal starting from an arbitrary system of generators. It can be seen as a generalization of the familiar Euclidean algorithm for polynomials in one variable and is comparable with the latter in its importance. It turns out that Buchberger's algorithm can be extended to the case of the noncommutative algebra D.

One can draw parallels between the concept of a Gröbner basis for an ideal $I \in D$ and E. Cartan's concept of a system of differential equations in involution. Similarly, Buchberger's algorithm for D can be compared with Cartan's procedure of "prolongation" of a given system $\{P_i\, u = 0\}$ by adding a sufficient number of differential consequences, i.e., equations of the form $R\,u = 0,\ R \in I$. However, in the approach of the classical period there seems to be no counterpart for the concept of a *reduced Gröbner basis*, which involves conditions preventing one from including redundant elements. In fact, in Version 2 above the reduced Gröbner basis is unique. This fact (together with the Buchberger algorithm) is the cornerstone of the applications of Gröbner bases to differential equations.

3. Hypergeometric functions and toric varieties

The point of view of computational algebra (that, for example, a polynomial $f(x), x = (x_1, ..., x_n)$ is regarded as a list of its monomials with coefficients) needs to be reconciled with that of geometry, where $f(x)$ is regarded as a function on $\mathbf{C}^n$. Speaking about individual monomials becomes natural if we consider $\mathbf{C}^n$ not as an abstract algebraic variety but as a variety with the action of the torus $(\mathbf{C}^*)^n$: then the monomials are the eigenfunctions of the action. So the type of algebraic geometry naturally adapted for symbolic computations is toric geometry studying toric varieties (equivariant partial compactifications of $(\mathbf{C}^*)^n$) and torus actions on more general varieties; cf. [2].

The monomial point of view being so natural by itself, many constructions of toric geometry have mathematical significance independently of any computational applications. For example, consider a linear action of $(\mathbf{C}^*)^n$ on a finite-dimensional vector space V. Such an action is diagonalizable: $V = \bigoplus_{\omega \in \mathbf{Z}^n} V_\omega$, where $x \in (\mathbf{C}^*)^n$ acts on V_ω via the Laurent monomial x^ω. Assume that the image of $(\mathbf{C}^*)^n \to \mathrm{Aut}(V)$ contains scalar operators and $\dim(V_\omega) = 0$ or 1. Let $\mathcal{A}$ be the set of ω such that $\dim(V_\omega) = 1$. The $\mathcal{A}$-hypergeometric system is a holonomic system of PDE on a function $\Phi(a), a \in V^*$. It depends on the choice of an infinitesimal character $\beta : \mathbf{C}^n = \mathrm{Lie}(\mathbf{C}^*)^n \to \mathbf{C}$ and (in an inessential way) on the choice of a generic $(\mathbf{C}^*)^n$-orbit $O \in V$. For any polynomial function F on V let F^∂ be the corresponding linear differential operator on V^* with constant coefficients. Then the system is:

$$F^\partial \Phi = 0, \quad \text{for} \quad \text{any} \quad F \quad \text{s.t.} \quad F|_O = 0; \tag{4}$$

$$L_\xi \Phi = \beta(\xi) \cdot \Phi, \quad \forall \xi \in \mathbf{C}^n. \tag{5}$$

Here L_ξ is the vector field on V^* given by the infinitesimal action of ξ.

This system was introduced by I.M. Gelfand, M.I. Graev and A.V. Zelevinsky [5], and its relations with toric varieties and algebraic geometry were found in [6], [7]. Notice that (4) intuitively means that $\widehat{\Phi}$, the Fourier transform of Φ, is a distribution on V supported on the closure $\overline{O}$, which, being a toric variety, has finitely many torus orbits, while (5) means that $\widehat{\Phi}$ is quasihomogeneous with respect to the torus action. Hence the typical possibility (in fact, the only one if β is generic) for $\widehat{\Phi}$ is $\widehat{\Phi} = \delta_O \cdot M$, where δ_O is the delta function of O and M is an appropriate monomial with complex exponents. So we can get solutions of (4-5) by taking "the" inverse Fourier transform of $\delta_O \cdot M$ (making sense of it in the complex domain involves the choice of a contour of integration, which is not unique). This leads to integral representations of solutions of (4-5) of algebro-geometric nature and, in particular, to the following conclusion: *For any fixed $\alpha_i, \gamma_j \in \mathbf{C}$ the integral*

$$I(f_1, ..., f_l) = \int_\sigma \prod_{i=1}^{l} f_i(t_1, ..., t_m)^{\alpha_i} t_1^{\gamma_1} ... t_m^{\gamma_m} dt_1 ... dt_m, \tag{6}$$

considered as a function of the indeterminate coefficients of the Laurent polynomials f_i, satisfies an appropriate $\mathcal{A}$-hypergeometric system. Here σ is understood to be a cycle in the local system determined by the f_i, α_i, γ_j and to vary with the f_i via the Gauss-Manin connection. Among the $I(f_1, ..., f_l)$ we find, for instance, periods of regular volume forms on complete intersections in toric varieties, which are of great interest in mirror symmetry; see [8].

It was shown in [7] that under mild genericity assumptions there is a basis in the space of solutions of (4-5) formed by explicit power series (called Γ-series) with each coefficient being a product of some values of the Gamma function. Intuitively, to arrive at a Γ-series, one can expand δ_O into a Laurent series (similarly to the identity $\sum_{n=-\infty}^{\infty} z^n = \delta_1(z)$ in one variable) and then perform the Fourier transform term by term.

These three aspects of the theory of $\mathcal{A}$-hypergeometric functions (the holonomic system, the integral representations and the Γ-series expansions) generalize the three ways of representing the famous hypergeometric function $F_{a,b,c}(z)$ of Gauss.

In fact, various classical generalizations of the Gauss function (of Pochhammer, Appel, Lauricella and others) are particular cases of the $\mathcal{A}$-hypergeometric formalism.

4. The book

The book interweaves the above three lines of thought in a very interesting way. First, and quite importantly, it provides a down to earth introduction to the formalism of D-modules. Not so long ago this whole area was regarded as very abstract. Accordingly, the examples likely to be discussed in the literature were mostly obtained by applying functorial operations like $f_*, f_!$ rather than by writing explicit systems of PDE. Chapter 1 of the book provides a different type of exposition, preserving the analytic flavor present in the theory of linear ODE. This chapter can be used as a first reading on the subject. The emphasis on explicit constructions will be much appreciated by people with backgrounds other than in abstract (motivic-style) algebraic geometry. Here is one example. It is clear theoretically that a holonomic system of linear PDE of any order with one unknown scalar function $u(x)$ can be replaced (near a generic point) by an equivalent system of first order equations on μ functions, where μ is the holonomic rank. In one variable, if we have an ODE of order μ, all one has to do is to consider the derivatives $u, u', ..., u^{\mu-1}$ as unknown functions. For several variables there is a similar completely explicit procedure, but it involves (not so surprisingly) Gröbner bases. This is explained clearly on pp. 38-39.

Already the introductory Chapter 1 (which serves also as a synopsis of all the main themes of the book) contains several examples of explicit holonomic systems, such as the system for integrals of products of linear forms [3], [4] and the general $\mathcal{A}$-hypergeometric system. Of course, an expert will recognize f_* or $f_!$ in any discussion of integrals, but a more explicit treatment will be beneficial for beginning students. An important contribution of the book is a higher-dimensional generalization of the classical method of Frobenius for finding series solutions to linear ODE with regular singularities. In this method, the series are formed out of monomials of the form $x^\alpha(\log x)^m$, $\alpha \in \mathbf{C}, m \in \mathbf{Z}$. One first finds possible "minimal terms" for such series and then develops each such term to a series by solving difference equations.

The higher-dimensional algorithm proceeds in a similar fashion, but with several nontrivial ingredients involved:

(A) The series in n variables that are considered can run in various directions, but each one has support (the set of exponents that actually enter into the series) lying in some convex cone. These cones are described by the Gröbner basis techniques. This point of view is different from the traditional treatment which basically sticks with more familiar Taylor series but has to appeal to the Hironaka desingularization theorem.

(B) The minimal term is also found using Gröbner-type techniques, namely by deforming the system to a torus invariant one, so that all the differential operators become polynomials in $x_i\partial/\partial x_i$. Such systems are just multiplicative analogs of systems of PDE with constant coefficients such as (1) and are solved in a similar way.

(C) All known definitions of the class of regular holonomic D-modules in higher dimensions are not very elementary. The authors choose the one with restrictions to arbitrary curves, and this necessitates the discussion of inverse images of

D-modules. Despite this higher level of abstraction, the authors bring their considerations to the level of explicit algorithms in §§2.5-2.6.

The authors then apply their algorithms to $\mathcal{A}$-hypergeometric systems and demonstrate how one can obtain Γ-series in this way. Since the derivation of Γ-series via the termswise Fourier transform of the Laurent expansion of δ_O is quite different in nature from the philosophy of Gröbner bases, it is in fact very interesting that general algorithms of Chapter 2 indeed give this kind of series. Further, the general approach covers as well the cases when there are not enough Γ-series to generate the space of solutions. In such cases (which are important for algebraic geometry) one has also to consider derivatives of the Γ-series with respect to the parameter β, and they involve logarithms as well as pure powers of variables [9], [10]. Again, the general algorithms applied to such situations produce these deformed series.

Along with the "applied" direction of constructing algorithms, the book contains the results of "purely" mathematical research by the authors on $\mathcal{A}$-hypergeometric systems, most importantly on the problem of determining the holonomic rank. In the majority of cases this rank is equal to the degree of the toric variety $\overline{O}$ which can be found as the volume of the convex hull of $\mathcal{A}$ (the Newton polytope). For example, this is true if β is generic enough. However, this is not always the case, and a counterexample was found by B. Sturmfels and N. Takayama [12]. In fact, the authors prove, as a by-product of their study of series solutions in Chapter 3, that the holonomic rank is always *greater or equal to* the volume. This may seem surprising from the point of view of general intuition about systems of PDE. For example, already two equations on a function of $n \geq 2$ variables, if chosen at random, will likely have no nonzero solutions at all, so the holonomic rank is 0. However, if we think in terms of systems depending on parameters, then such "unlucky choices" of equations correspond to more generic and not more special values of the parameters, so it is in fact natural to expect that the dimension of the space of solutions will jump and not drop for special values. In Chapter 4 the authors study the relation between the rank and the volume in a systematic way. Since for generic β there is equality, one is led naturally to studying the locus of "exceptional" β, for which a jump occurs. It is known that this locus lies in the union of certain "resonant" hyperplanes, but its more precise structure is not known. A more familiar instance of such resonant behavior in the space of parameters can be seen in the theory of the b-function (the Bernstein-Sato polynomial) $b(s)$ associated to a polynomial $f(x_1, ..., x_n)$ which controls the holonomic system for the multivalued function $f^s, s \in \mathbf{C}$.

The most technically advanced and impressive part of the book is Chapter 5, which provides a bridge between functorial operations on D-modules and the computational approach adopted in the book. Besides a treatment of b-functions from the Gröbner basis point of view, the authors explain how one can, in principle, compute inverse and direct images of D-modules, following the work of T. Oaku, N. Takayama and other people. A striking example is the direct image of the constant D-module from an open set U in $\mathbf{C}^n$ obtained as a complement to a hypersurface Z. This leads to an algorithm for computing topological (de Rham) cohomology of U and Z (note that Z is allowed to be singular). The input of such an algorithm is just the equation of Z (a list of monomials with coefficients), and the output is the list of the ranks of the cohomology groups. These algorithms have in fact been implemented in computer systems described in the Appendix, and anyone can try them out in practice. In a similar way, various determinations of integrals such as (6) are labeled by topological cohomology with coefficients in a local system and also fall within the framework of this chapter.

Algebraic geometers already got used to the possibility of computing the cohomology of coherent sheaves by hitting some keys on a keyboard, but questions pertaining to the topological cohomology of algebraic varieties (especially open or singular) are still regarded as belonging to some kind of "Alaskan refuge" of contemplative thinking. The possibility of an "industrial approach" to this area has a sobering effect, forcing one to ponder wider issues raised by the use of computers.

In summary, I think that Saito, Sturmfels and Takayama have written an extremely useful and timely text which will appeal to people with different backgrounds and different purposes in mind. I can think of at least four categories of such people.

First, beginners wishing to understand D-modules may use about one-third of the book to see how the general theory develops the familiar ideas from ODE. This category would include not just students but also more seasoned analysts who might not have other occasions to see why D-modules are relevant to the "true" theory of differential equations. In any event, a part of the book can be used as a basis for an introductory, example-oriented course on D-modules. Second, experts on D-modules will find here a refreshing alternative to the motivic or representation-theoretic points of view and new food for thought. Third, people interested in hypergeometric functions and their numerous applications (such as mirror symmetry) will no doubt find the book a useful reference. Finally, the primary intended readership of the book consists of people working on computational questions. The authors have summarized a period of rapid development of the computational theory of differential equations, and this well written book is necessary reading for anyone interested in this area.

References

[1] D. Cox, J. Little, D. O'Shea, Ideals, Varieties and Algorithms, Springer, 1992. MR **93j:**13031

[2] W. Fulton, Introduction to Toric Varieties, Princeton Univ. Press, 1993. MR **94g:**14028

[3] I.M. Gelfand, General theory of hypergeometric functions, *Sov. Math. Dokl.* **33** (1986), 573-577. MR **87h:**22012

[4] I.M. Gelfand, S.I. Gelfand, Generalized hypergeometric equations, *Sov. Math. Dokl.* **33** (1986), 643-646. MR **87h:**22013

[5] I.M. Gelfand, M.I. Graev, A.V. Zelevinsky, Holonomic systems of equations and series of hypergeometric type, *Soviet Math. Dokl.* **36** (1988), 5-10. MR **88j:**58118

[6] I.M. Gelfand, M.M. Kapranov, A.V. Zelevinsky, Generalized Euler integrals and A-hypergeometric functions, *Adv. Math.* **84** (1990), 255-271. MR **92e:**33015

[7] I.M. Gelfand, A.V. Zelevinsky, M.M. Kapranov, Hypergeometric functions and toric varieties, *Funct. Anal. Appl.* **23** (1989), 94-106; see also correction, *ibid.* **27** (1993), 295. MR **90m:**22025; MR **95a:**22010

[8] S. Hosono, B. Lian, S.-T. Yau, GKZ-generalized hypergeometric functions in mirror symmetry of Calabi-Yau hypersurfaces, *Comm. Math. Phys.* **182** (1996), 535-577. MR **98g:**14042

[9] S. Hosono, B. Lian, S.-T. Yau, Maximal degeneracy points of GKZ systems, *Journal of the AMS* **10** (1997), 427-443. MR **97j:**32014

[10] J. Stienstra, Resonant hypergeometric systems and mirror symmetry, in: "Integrable Systems and Algebraic Geometry" (Proceedings of the Taniguchi Symposium 1997, M.-H. Saito, Y. Shimizu, K. Ueno, Eds.), World Scientific, 1998. MR **2000b:**14013

[11] B. Sturmfels, Gröbner Bases and Convex Polytopes, Amer. Math. Soc., 1996. MR **97b:**13034
[12] B. Sturmfels, N. Takayama, Gröbner bases and hypergeometric functions, in: "Gröbner Bases and Applications" (Linz, 1998), pp. 246-258, London Math. Soc. Lect. Series **251**, Cambridge Univ. Press, 1998. MR **2001c:**33026

Mikhail Kapranov
University of Toronto
E-mail address: kapranov@math.toronto.edu