Ecuaciones Lineales sobre Anillos de Valuación

Joachim von zur Gathen
Department of Computer Science
University of Toronto

Resumen

Se presentan algoritmos de tiempo polinomial para resolver sistemas de ecuaciones lineales sobre $R[x]$, cuando R es un anillo con valuación euclidiana.

1. Introducción

Resolver sistemas de ecuaciones lineales es claramente uno de los más importantes problemas de computación. También ha ganado aceptación la noción que "tiempo polinomial" - es decir un número total de operaciones que es polinomial en el tamaño de la entrada - es una condición necesaria para que un algoritmo sea practicable.

La eliminación de Gauss presenta un algoritmo para resolver sistemas de ecuaciones lineales en tiempo polinomial, de hecho en $O(n^3)$ operaciones para sistemas de $n \times n$ sobre un campo de base arbitrario. Pero, cuando los elementos del campo de base son dados por representaciones de tamaño variable - como por ejemplo la representación binaria de números enteros - no es evidente que los resultados intermedios en la eliminación de Gauss son de tamaño polinomial, y que la eliminación se puede efectuar en un número polinomial de operaciones binarias. Edmonds [1967] mostró que los resultados intermedios no son demasiado grandes en este caso especial.

En un espíritu similar, se conocen hoy en día algoritmos polinomiales para los siguientes problemas: solución de sistemas lineales sobre $\mathbb{Z}$ (von zur Gathen-Sieveking [1976]), calculación de la forma normal triangular "de Hermite" y de la forma diagonal "de Smith" para matrices sobre $\mathbb{Z}$ (Kannan-Bachem [1979]) y sobre $\mathbb{Q}[x]$ (Kannan [1983]).

En este trabajo, presentamos un algoritmo de tiempo polinomial para calcular las formas normales de Hermite y de Smith para matrices cuyos elementos son tomados de $R[x]$, donde R es un anillo con valuación euclidiana. Esto es una generalización simultánea de los casos $R=\mathbb{Z}$ y $R=F[y]$ para un campo F arbitrario, e inductivamente también para anillos de polinomios en muchas variables. La noción de "anillo euclidiano" es el concepto correcto para describir algoritmos como el cálculo del máximo común divisor y el algoritmo chino del resto; en nuestro caso, "anillo con valuación euclidiana" es una nueva noción que parece ser el concepto correcto para métodos de factorización de polinomios (von zur Gathen [1983]), y para las ecuaciones lineales de este trabajo.

2. Valuaciones

Por definición, una valuación $v: R \to \mathbb{R}$, donde R es un anillo (conmutativo, con 1) sin divisores de zero, cumple las propiedades siguientes para todo $a, b \in R$:

(i) $v(a) \geq 0$,

(ii) $v(a) = 0 \iff a = 0$,

(iii) $v(ab) = v(a)v(b)$,

(iv) $v(a+b) \leq v(a) + v(b)$.

v se llama no-arquimediana si

(iv)' $v(a+b) \leq max\{v(a), v(b)\}$.

Para propiedades elementales de valuaciones, ver por ejemplo van der Waerden [1970]. Los dos ejemplos importantes para el punto de vista de este trabajo son: la "valuación absoluta" sobre $R = \mathbb{Z}$ con $v(a) = |a|$; y la "valuación del grado" v sobre $R = F[x]$ con $v(f) = 2^{\text{grad}f}$ (aunque es más común considerar la "valuación logarítmica" $w(f) = \text{grad}\, f$), y valuaciones similares del grado sobre $R = F[x_1,\ldots,x_k]$.

Fijamos una valuación v sobre R. Escribimos $M_{m\times n} = M_{m\times n}(R[x])$ para el conjunto de matrices de $m \times n$ cuyos elementos provienen de $R[x]$, y F para el campo de cocientes de R. $\text{grad}f$ es el grado de un polinomio $f \in R[x]$. Si $A \in M_{m\times n}$ y $b \in M_{m\times 1}$, entonces escribimos $(A\,|\,b) \in M_{m\times(n+1)}$ para la matriz aumentada. Utilizaremos la definición siguiente.

Definición 2.1. Sean $m, n \geq 1$, y $A \in M_{m\times n}$. Entonces

$$\text{grad}A = \max_{i,j} \text{grad}A_{ij},$$

$$v(A) = \max_{i,j} v(A_{ij}).$$

$$v_2(A) = (\sum_{i,j} v(A_{ij})^2)^{1/2}. \quad \square$$

Identificamos un polinomio $f \in R[x]$ con el vector de sus coeficientes, que es miembro de $M_{n\times 1}$ para $\text{grad} f < n$, y utilizaremos $v(f)$ como definida más arriba. Para una matriz $A \in Mmn$, usamos $A_{i*} \in M_{1\times n}$ para denotar la i-ésima fila de A.

Lema 2.2 (Desigualdad de Hadamard) Sea $A \in M_{n\times n}$. Entonces

$$\text{grad}\det A \le n\ \text{grad}A,$$

$$v(det\ A) \le \prod_{1\le i\le n} v_2(A_{i*}) \le (n^{1/2}v(A))^n.$$

Si v es no-arquimediana, entonces

$$v(det\ A) \le v(A)^n. \quad \square$$

Lema 2.3. Sea $A \in M_{m\times n}$, $b \in M_{m\times 1}$, $d = \text{grad}(A\,|\,b)$, $w = v(A\,|\,b)$, y supongamos que existe $y \in F[x]^n$ tal que $Ay = b$. Entonces existen $y_1,\dots,y_n \in R[x]$ y $y_0 \in R\setminus\{0\}$ tal que para todo i valen

$$A(y_1/y_0,\dots,y_n/y_0) = b\ ,$$

$$\text{grad}\, y_i \le 2md\ ,$$

$$v(y_i) \le (m\sqrt{d+1}w)^{3m^2(d+1)}.$$

Demostración. Podemos suponer que el rango de A es igual a m, y que las primeras m columnas de A son linealmente independientes (sobre $F(x)$). Sea $B \in M_{m\times m}$ la matriz formada por las primeras m columnas, y $C \in M_{m\times(n-m)}$ formada por las últimas columnas de A.

Tomamos $t_1,\dots,t_n \in F[x]$ tal que

$$A(t_1,\dots,t_n) = b\ .$$

Sea $u = \det B \in R[x]$ y $q_i, r_i \in F[x]$ para $1\le i\le n$ tal que

$$t_i = q_i u + r_i\ ,$$

$$\text{grad}r_i < \text{grad}u \le m\ gradB \le md\ ,$$

$$r = (r_{m+1},\dots,r_n) \in R[x]^{n-m}\ ,$$

$$(z_1,\dots,z_m) = (t_1,\dots,t_m) + \text{adj}(B)\cdot C\cdot(q_{m+1},\dots,q_n) \in F[x]^m\ ,$$

donde $\text{adj}(B) \in M_{m\times m}$ es la adjunta de B. Ahora el sistema no-singular de ecuaciones

lineales

$$Bz = b - Cr$$

tiene la única solución

$$z = (z_1,\ldots,z_m)$$

en $F(x)^m$. Utilizando la regla de Cramer, se obtiene que cada $\det B \cdot z_i \in R[x]$ es un subdeterminante de $(B \mid b - Cr)$ y tiene un grado menor que $(m-1)d + (m+1)d = 2md$, porque $\operatorname{grad}(b-Cr) < d + md = (m+1)d$. Resulta que $\operatorname{grad} z_i < 2md$ para $1 \le i \le m$.

Ahora consideramos $2nmd$ indeterminadas s_{ij} sobre F, para $1 \le i \le n$ y $0 \le j < 2md$, y el sistema de $m(2m+1)(d+1)$ ecuaciones lineales con coeficientes de R que resulta de la condición

$$A \cdot (\sum_j s_{1j}x^j, \ldots, \sum_j s_{nj}x^j) = b$$

al comparar coeficientes de potencias de x. Por lo supuesto más arriba, este sistema tiene una solución en F. Sigue de la regla de Cramer y la desigualdad de Hadamard (Lema 2.2) que también tiene una solución $(y_{ij}/y)_{i,j}$ con $y_{ij}, y \in R$ y

$$v(y_{ij}), v(y) \le ((m(2m+1)(d+1))^{1/2}w)^{m(2m+1)(d+1)}$$

$$\le (m\sqrt{d+1}w)^{3m^2d}. \quad \square$$

Sean $f, g \in R[x]$, y $h \in F[x]$ el único máximo común divisor ("mcd") mónico, es decir con coeficiente máximo igual a 1. La teoría de subresultantes (Collins [1967], Brown [1971]; ver también Borodin-von zur Gathen-Hopcroft [1982]) da un método de calcular h como la única solución de un sistema de ecuaciones lineales con coeficientes tomados de R. Si p es el determinante de esta matriz de coeficientes (la "subresultante principal"), entonces $ph \in R[x]$ y existen $s, t \in R[x]$ tal que

$$sf + tg = ph.$$

En este trabajo, usamos $\operatorname{mcd}_R(f, g)$ para denotar este $ph \in R[x]$.

Lema 2.4. Sean $f, g \in R[x]$ de grado no mayor que n, y $v(f), v(g) \le w$. Entonces existen $k, s, t \in R[x]$ tal que

$$k = \operatorname{mcd}_R(f, g) = sf + tg$$

$$\operatorname{grad} k, \operatorname{grad} s, \operatorname{grad} t \le n$$

$$v(k), v(s), v(t) \le (2nw)^{3n^2}$$

Se pueden calcular k, s, t resolviendo un sistema no singular de ecuaciones lineales sobre R.

Demonstración. Es claro por las observaciones precedentes y Lema 2.3. □

El lema siguiente proviene de Mignotte [1982].

Lema 2.5 (Landau-Mignotte) Supongamos que $f, g \in F[x]$ son mónicos, $m = \text{grad} g$, y que g es un divisor de f. Entonces

$$v(g) \leq v(2^m) v_2(f).$$

Si v es non-arquimediana, entonces

$$v(g) \leq v(f). \quad \square$$

3. Triangulación parcial

Llamamos una matriz $U \in M_{m \times m}(F[x])$ *unimodular* si $detU \in F$. El algoritmo FORMA TRIANGULAR, que se presentará en la próxima sección, transforma una matriz $A \in M_{m \times n}$ a forma triangular superior por operaciones unimodulares, es decir el algoritmo calcula una matriz triangular $B \in M_{m \times n}$ y una matriz unimodular $U \in M_{m \times m}$ tal que $A = UB$.

Como en la eliminación de Gauss, el algoritmo va creando submatrices cuadradas de forma triangular más y más grandes en el ángulo izquierdo superior de la matriz considerada. En esta sección, presentamos el algoritmo básico que agrega una fila y una columna nuevas al cuadrado.

Algoritmo FORMA TRIANGULAR PARCIAL

Entrada: $B \in M_{k \times k}(R[x])$, con el menor de $(k-1)\times(k-1)$ de B de forma triangular superior, y $\det B \neq 0$.

Salida: $C, U \in M_{k \times k}(R[x])$, con $B = UC$, C de forma triangular superior, y $\det U = 1$.

(1) Poner $C_0 = B$, $U_0 = I_k$, y calcular $C_i, U_i, V_i \in M_{k \times k}$ para $i = 1,\ldots,k-1$ en los pasos 2 y 3.

(2) Poner $d_i = (C_{i-1})_{ii}$, $e_i = (C_{i-1})_{ki}$, y calcular $g_i, s_i, t_i \in R[x]$ tal que $g_i = \text{mcd}_R(d_i, e_i) = s_i d_i + t_i e_i$.

(3) Poner

$$V_i = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & s_i & \cdots & t_i \\ & & \vdots & \ddots & \vdots \\ & & -e_i/g_i & \cdots & d_i/g_i \end{bmatrix},$$

$$C_i = V_i C_{i-1},$$

$$U_i = U_{i-1} V_i^{-1}.$$

(4) Devolver $C = C_{k-1}$ y $U = U_{k-1}$.

Lema 3.1. El algoritmo FORMA TRIANGULAR PARCIAL produce los resultados descriptos más arriba. Se puede efectuar con $O(k^5 d^4)$ operaciones en R, utilizando solamente elementos $r \in R$ con

$$v(r) \leq c^d (k v(B))^{k^2},$$

donde $d = \text{grad} B$ y c es una constante universal.

Demostración: Por inducción sobre i se demuestra fácilmente que los resultados producidos son correctos, es decir que $B = UC$, C es de forma triangular superior, y $\det U = 1$. La única parte no trivial es la cota superior sobre el tamaño de los resultados intermedios.

Para cada valor de i, los pasos 2 y 3 solamente cambian las filas i y k de C_{i-1}. En particular, $(C_{i-1})_{i*} = B_{i*}$. Para $1 \leq i, j \leq k$ sea

$$f_{ij} = \det \begin{bmatrix} B_{11} & \cdots & B_{1i} & B_{1j} \\ 0 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & B_{ii} & B_{ij} \\ B_{k1} & \cdots & B_{ki} & B_{kj} \end{bmatrix}.$$

Vamos a demostrar que para $0 \leq i < k$ y $1 \leq j \leq k$ vale lo siguiente:

$$(C_i)_{kj} = (g_1 \cdots g_i)^{-1} f_{ij}.$$

La demostración se efectúa por inducción sobre i, el caso $i = 0$ siendo claro. Por $0 < i < k$ tenemos

$$(C_i)_{kj} = -\frac{e_i}{g_i}(C_{i-1})_{ij} + \frac{d_i}{g_i}(C_{i-1})_{kj}$$

$$= g_i^{-1}(-(C_{i-1})_{ki}B_{ij} + B_{ii}(C_{i-1})_{kj})$$

$$= (g_1 \cdots g_i)^{-1}(-f_{i-1,j}B_{ij} + B_{ii}f_{i-1,j}) = (g_1 \cdots g_i)^{-1}f_{ij},$$

donde la última ecuación sigue por expansión de la matriz para f_{ij} según su i-ésima fila.

También se verifica fácilmente que

$$V_i^{-1} = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & d_i/g_i & \cdots & -t_i & \\ & & & \cdot & 1 & \cdot & \\ & & & \cdot & & \ddots & \cdot \\ & & & \cdot & & & 1 \ \cdot \\ & & & e_i/g_i & \cdots & s_i & \end{bmatrix}$$

$$(U_i)_{j*} = (V_j^{-1})_{j*} \text{ por } i \geq j,$$

$$(U_i)_{k*} = (e_1/g_1, d_1e_2/g_1g_2, \ldots,$$

$$d_1 \cdots d_{i-1}e_i/g_1 \cdots g_i, 0, \ldots, 0, d_1 \cdots d_i/g_1 \cdots g_i)$$

Las cotas siguientes son claras ahora, para todo valor h, i, j, $1 \leq h, i, j \leq k$:

$$\text{grad}(f_{ij}) \leq kd,$$

$$\text{grad}(C_i)_{hj} \leq kd,$$

$$\text{grad}(U_i)_{hj} \leq k^2d,$$

$$\text{grad}(g_i), \text{grad}(s_i), \text{grad}(t_i) \leq kd,$$

$$v(f_{ij}) \leq (k^{1/2}w)^k,$$

$$v((C_i)_{hj}) \leq v(2^{kd})v_2(f_{ij}) \leq v(2^{kd})\sqrt{kd}\,(k^{1/2}w)^k,$$

$$v(d_1 \cdots d_i) \leq (k^2d)^{\log k}(v(2^{kd})\sqrt{kd}\,(k^{1/2}w)^k)^k \leq (kw)^{k^2}c^d,$$

$$v(U_i)_{hj} \leq (kw)^{k^2}c^d,$$

con una constante universal c. El costo dominante del cómputo ocurre en el paso 2.

En éste se calcula el mcd de polinomios de grado no mayor que kd, y cada mcd se puede calcular con $O((kd)^4)$ operaciones en R. Así, el costo total es $O(k^5d^4)$. Para cada resultado intermedio $r \in R$ vale

$$v(r) \leq c^d (kw)^{k^2}$$

para una constante c. □

4. Forma normal de Hermite y de Smith

En esta sección, calculamos dos formas normales para matrices, las de Hermite y de Smith. Llamamos a una matriz $A \in M_{m \times n}$ *principalmente no-singular* si todos los menores principales (= submatrices cuadradas en el ángulo izquierdo superior) de A son no-singulares. Consideraremos solo matrices de este tipo por simplicidad: nuestro algoritmo, como la eliminación de Gauss, no necesita ninguna permutación de columnas para estas matrices. Pero no es una restricción severa, porque una matriz general de rango maximal se transforma fácilmente a una matriz principalmente singular. Una variación del algoritmo también se aplica al caso de matrices que no tienen rango maximal.

Definición 4.1. Una valuación $v: R - \{0\} \to \mathbb{N}$ se llama una valuación euclidiana si existe β, $0 < \beta < 1$, tal que

$$E_1\colon \forall a \in R \quad (a \neq 0 \Rightarrow v(a) \geq 1)$$

$$E_2\colon \forall a, b \in R \ \exists q \in R \quad (b \neq 0 \Rightarrow v(a - qb) \leq \beta v(b)).$$

En este caso, R se llama un anillo con valuación euclidiana. □

La condición E_2 dice que la división con resto es posible, con un resto de valor no más de β veces el valor del divisor. Dicho anillo entonces es euclidiano (en el sentido común), y el algoritmo euclidiano para calcular un máximo común divisor de $a, b \in R$ se puede efectuar con no más de $1 + \frac{\log(v(b))}{\log(1/\beta)} = O(\log v(b))$ divisiones. Tenemos dos ejemplos importantes de anillos con valuación euclidiana: $\mathbb{Z}$ con el valor absoluto, y $F[x]$ con $v(f) = 2^{grad f}$, donde F es un campo. En ambos casos podemos poner $\beta = 1/2$.

Observación 4.2. También podríamos definir un anillo con "valuación pseudo-euclidiana", en el cual se puede efectuar la "pseudo-división". Es decir, se substituye la condición E_2 por

$$\forall a,b \in R \ \exists c,q \in R \quad b \neq 0 \Rightarrow v(c) = 1 \text{ y } v(ca - qb) \leq \beta v(b).$$

Si F es un anillo sin divisores de zero, esta condición será satisfecha por $F[x]$ con $v(f) = 2^{grad f}$.

Un polinomio $f = \sum f_i x^i \in R[x]$ se llama *primitivo* si $\text{mcd}(f_0,...,f_n) \in R$ es invertible. Para cada $g \in F[x]$ se pueden calcular $a \in F$ y un polinomio primitivo $f \in R[x]$ tal que $g = af$; tales a y f son únicos módulo multiplicación por un elemento invertible de R.

Algoritmo FORMA TRIANGULAR.

Entrada: $A \in M_{n \times n}(R[x])$ principalmente no-singular, donde R es un anillo con valuación euclidiana.

Salida: $D \in M_{n \times n}(F[x])$ de forma triangular superior, y $V \in M_{n \times n}(F[x])$ unimodular tal que $A = VD$.

(1) Poner $A_0 = A$, $V_0 = I_n$, $d = \text{grad} A$. Por $k = 1,...,n-1$ calcular A_k, $V_k \in M_{n \times n}$ en los pasos 2 hasta 5.

(2) Llamar el algoritmo FORMA TRIANGULAR PARCIAL con el $k \times k$-menor de A_{k-1} como entrada, y $B_k \in M_{k \times k}(R[x])$ como salida.

(3) Calcular polinomios primitivos $b_1,...,b_k \in R[x]$ tal que el elemento diagonal i-ésimo de B_k es un múltiplo escalar de b_i, con el escalar tomado de F.

(4) Considerar el sistema de $k(k+1)/2$ ecuaciones lineales sobre $R[x]$ (en las k^2 entradas indeterminadas W_{ij} de $W = (W_{ij})_{i,j}$) que corresponde a la condición

$$W_k P_k = \begin{bmatrix} b_1 & & & \\ 0 & \cdot & & * \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \\ 0 & \cdots & 0 & b_k \end{bmatrix}.$$

donde P_k es el principal menor de $k \times k$ de A, y no se impone ninguna condición a la parte triangular estrictamente superior. Calcular una solución $W = (w_{ij}/w_{oo})_{i,j}$ con

$$w_{ij} \in R[x], \ w_{00} \in R \setminus \{0\},$$

$$\text{grad} w_{ij} \leq 2k^2 d$$

$$v(w_{ij}) \leq (\sqrt{d}\, v(2^d) v(A))^{4k^3 d},$$

para todo i, j, donde $d = \text{grad}A$.

(5) Poner

$$T = \begin{bmatrix} W_k & & 0 \\ & 1 & \\ & & \ddots \\ 0 & & \\ & & 1 \end{bmatrix} \in M_{n\times n}(R[x]),$$

$A_k = TA$, y $V_k = T^{-1}$.

(6) Devolver $D = A_{n-1}$ y $V = V_{n-1}$.

Teorema 4.3. Sea $A \in M_{n\times n}(R[x])$ principalmente no-singular, y $d = \text{grad}A$, $w = v(A)$. El algoritmo FORMA TRIANGULAR se puede efectuar y produce los resultados descriptos más arriba. El número de operaciones en R es $(nd)^{O(1)}$, y todo resultado intermedio r tiene tamaño polinomial, es decir

$$v(r) = w^{(nd)^{O(1)}}. \quad \square$$

Corolario 4.4. Si R es un anillo con valuación euclidiana, se puede calcular la forma normal de Hermite para matrices en $R[x]$ en tiempo polinomial. $\square$

Como mencionado en la introducción, los casos especiales $R = \mathbf{Z}$ (Kannan-Bachem [1979]) y $R = \mathbf{Z}[x]$ o $R = \mathbf{Q}[x]$ (Kannan [1983]) son bien conocidos. El resultado de este trabajo se aplica también a los casos $R = F[x]$ y $R = F[x, y]$ para un campo F arbitrario, cuando una operación en F tiene un costo constante. Esto es realista en el caso de un campo finito, por ejemplo.

Ahora es fácil aplicar el método también a la traspuesta de una matriz ya en forma normal de Hermite, y se obtiene el resultado siguiente.

Corolario 4.5. Si R es un anillo con valuación euclidiana, se puede calcular la forma normal diagonal de Smith de una matriz de $R[x]$ en tiempo polinomial. $\square$

Corolario 4.6. Si R es un anillo con valuación euclidiana, se puede - en tiempo polinomial - decidir si un sistema de ecuaciones lineales tiene una solución, y en caso afirmativo, calcular una solución. $\square$

Bibliografía

A. Borodin, J. von zur Gathen, and J. Hopcroft, Fast parallel matrix and gcd computations. Information and Control **52** (1982), 241-256.

W.S. Brown, On Euclid's algorithm and the computation of polynomial Greatest Common Divisors. J. ACM **18** (1971), 478-504.

G.E. Collins, Subresultants and Reduced Polynomial Remainder Sequences. J. ACM **14** (1967), 128-142.

J. Edmonds, Systems of Distinct Representatives and Linear Algebra. J. of Res. Nat. Bureau of Standards **71B** (1967), 241-245.

J. von zur Gathen and M. Sieveking, Weitere zum Erfüllungsproblem polynomial äquivalente kombinatorische Aufgaben, in: Komplexität von Entscheidungsproblemen, ed. by E. Specker and V. Strassen, Lecture Notes in Computer Science 43 (1976), 49-71.

J. von zur Gathen, Computations in rings with valuations. Proc. 3rd Conf. Foundations of Software Technology and Theoretical Computer Science, Bangalore, 1983, 111-128. To appear in Math. Comp.

R. Kannan and A. Bachem, Polynomial algorithms for computing the Smith and Hermit normal forms of an integer matrix. SIAM J. Comput. **8** (1979), 499-507.

R. Kannan, Solving systems of linear equations over polynomials. Proc. 3rd Conf. Foundations of Software Technology and Theoretical Computer Science, Bangalore, 1983, 129-144.

M. Mignotte, Some useful bounds. In: Computer algebra, symbolic and algebraic computation, ed. B. Buchberger, G.E. Collins and R. Loos, Computing, Supplementum 4, Springer Verlag, 1982, 259-263.

B.L. van der Waerden, Modern Algebra, vol. 1. Ungar, New York, 1970.