PUBLICATIONS MATHÉMATIQUES ET INFORMATIQUES DE RENNES

ANNETTE PAUGAM

Algorithmes d'élimination des quantificateurs

Publications de l'Institut de recherche mathématiques de Rennes, 1985, fascicule 4
« Séminaires de mathématiques - science, histoire et société », , p. 173-195

<http://www.numdam.org/item?id=PSMIR_1985___4_173_0>

ALGORITHMES D'ELIMINATION DES QUANTIFICATEURS

Par **Annette PAUGAM**

Le principe de l'élimination des quantificateurs sur les corps réels clos est un résultat classique (Tarski 1948) de la théorie des modèles. Le développement récent de la géométrie semi-algébrique et du calcul formel ont remis cette question d'actualité. Je m'attacherai dans cet exposé à la comparaison entre trois algorithmes d'élimination issus des travaux de Seidenberg [S], Hormander [H] et Collins [C].

Enonçons d'abord le théorème qui précise en quoi consiste l'élimination sur les corps réels clos. C'est le principe de Tarski-Seidenberg.

Théorème : Etant donnés un corps K ordonnable et $P_1 \ldots P_s$ une suite de polynômes dans $K[X_1 \ldots X_n, Y]$ et $P_1 \varepsilon_1 0, \ldots, P_s \varepsilon_s 0$ (ε_i désignant $<$, $>$ ou $=$) des conditions de signes sur les polynômes P_i, alors il existe une combinaison booléenne d'égalités et d'inégalités polynômiales en $\underline{X} = (X_1 \ldots X_n)$ à coefficients dans K : $\mathcal{S}(\underline{X})$ telle que, pour tout corps réel clos R contenant K et tout $\underline{x}$ dans R^n, l'existence d'un y de R tel que

$$(P_1(\underline{x},y)\varepsilon_1 0) \text{ et } (P_2(\underline{x},y)\varepsilon_2 0) \ldots \text{ et } (P_s(\underline{x},y)\varepsilon_s 0)$$

soit équivalente à $\mathcal{S}(\underline{x})$.

La démonstration de ce théorème figure dans plusieurs ouvrages ([Kr], [J],[BCR]). Dans cet énoncé y est "éliminé". On peut continuer le procedé

par récurrence et aboutir ainsi à l'élimination de plusieurs quantificateurs. On peut remarquer que les conditions obtenues sur $\underline{x}$ ne dépendent ni du corps réel clos R, ni du $\underline{x}$ choisi dans R^n. Une preuve algorithmique de ce théorème permet en théorie de résoudre les problèmes d'existence réelle "à la machine". La pratique, compte tenu de l'importance des calculs, ne sera peut être pas aussi simple. Nous le verrons sur l'exemple du problème suivant : trouver des conditions sur x_0, y_0, a et b pour que l'ellipse

$$E(x,y) = \frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} - 1 = 0$$

soit incluse dans le cercle

$$C(x,y) = x^2+y^2-1 = 0,$$

ce qui peut s'écrire

$$\text{non} \quad ((\exists x)(\exists y)\ a^2b^2E(x,y) = 0 \quad \text{et} \quad C(x,y) \geqslant 0)$$

C et a^2b^2E étant des polynômes de $\mathbb{Q}[x_0,y_0,a,b,x,y]$.

Ce problème a été posé par Kahan en 1975 [Ka]. Un cas particulier facile à été traité par Arnon [A], grâce à l'algorithme de Collins, il a été résolu par Lauër [Lau], Mignotte [M] et Lazard [Laz] hors algorithme.

Ce travail m'a été proposé par M.-F. Coste-Roy. Je l'ai réalisé dans le cadre du séminaire de géométrie algébrique réelle et du groupe de travail de calcul formel de Rennes, notamment en collaboration avec M.-F. Coste-Roy et I. Giorgiutti.

I - Présentation des trois algorithmes.

Nous utiliserons dans les trois algorithmes la notion de pseudo-division dans $K[X_1 \ldots X_n\ Y]$. Etant donnés deux polynômes $A(\underline{X},Y)$ et $B(\underline{X},Y)$, la division euclidienne dans $K(\underline{X})[Y]$ fait apparaître au dénominateur la puissance p du coefficient dominant $B_0(\underline{X})$ de B. La pseudo-division sera la division euclidienne de $B_0(\underline{X})^{2k}$ A par B dans $K(\underline{X})[Y]$ où 2k est le plus petit entier pair supérieur à p.

$$B_0(\underline{X})^{2k}A = B(\underline{X},Y)Q(\underline{X},Y) + R(\underline{X},Y) \quad \text{avec } \deg_Y R < \deg_Y B .$$

Compte tenu du choix de 2k, Q et R sont en fait des polynômes en $\underline{X}$ et Y appelés pseudo-quotient et pseudo-reste. De plus, pour $(\underline{x},y) \in R^{n+1}$, le signe de $A(\underline{x},y)$ quand $Q(\underline{x},y) = 0$ est le même que celui de $R(\underline{x},y)$.

Rappelons aussi le théorème de Sturm qui nous servira dans les algorithmes de Seidenberg et Collins.

Théorème : Soit P un polynôme à coefficients dans un corps réel clos et soit la suite $P_0,P_1,\ldots,P_s$ de polynômes définie par $P_0(x) = P(x)$, $P_1(x) = P'(x)$,

$$\begin{aligned}
&P_0(x) = Q_1(x)P_1(x) - P_2(x) \quad \text{avec} \quad \deg P_2 < \deg P_1 \\
&\vdots \\
&P_{i-1}(x) = Q_i(x)P_i(x) - P_{i+1}(x) \quad \text{avec} \quad \deg P_{i+1} < \deg P_i \\
&P_{s-1}(x) = Q_s(x)P_s(x) .
\end{aligned}$$

Supposons que [a,b] soit un intervalle tel que $P(a) \neq 0$ et $P(b) \neq 0$. Alors le nombre de zéros distincts de P dans [a,b] est $V_a - V_b$ où V_c désigne le nombre de variations de signe dans la suite $P_0(c),P_1(c)\ldots P_s(c)$.

La démonstration de ce théorème figure dans [J] et dans [BCR].

I_1 - Algorithme de Hormander.

On trouvera une démonstration détaillée de cet algorithme de Hormander [H] dans [BCR].

Soit $P_1 \ldots P_s$ une suite de polynômes de $K[X_1 \ldots X_n\ Y]$ et soient $\varepsilon_1 \ldots \varepsilon_s$ des signes $(>,<,=)$. Pour résoudre la question : quelles sont les conditions de signe sur $\underline{x}$ pour que

(*) il existe $y \in R$ tel que $P_1(\underline{x},y)\,\varepsilon_1 0$ et... et $P_s(\underline{x},y)\,\varepsilon_s 0$.

Nous allons construire des polynômes en $\underline{X} : B_1 \ldots B_m$ tels que lorsque les signes de $B_1(\underline{x}),\ldots,B_m(\underline{x})$ sont fixés, le "tableau de signe" des polynômes P_i est déterminé. Précisons qu'un tableau de signe d'une suite $P_1 \ldots P_s$ de polynômes en Y sera la donnée du nombre N des zéros $y_1 \ldots y_N$ des différents polynômes P_i et d'un tableau à s lignes et 2N+1 colonnes donnant le signe de chaque polynôme P_i en chaque zéro y_i et sur chaque intervalle $]y_i, y_{i+1}[$. Mais ce tableau ne donne pas la valeur des zéros.

La réponse positive ou négative à la question (*) ne dépend alors que du signe des différents B_i en $\underline{x}$ puisqu'elle est donnée par l'examen des colonnes du tableau de signe des P_i.

Etant donnés les polynômes $P_1 \ldots P_s$ de $K[X_1 \ldots X_n, Y]$ nous construisons, sous l'hypothèse que les coefficients dominants ne s'annulent pas, une suite de 2s polynômes en dérivant un des polynômes de plus haut degré en Y soit P_1, par rapport à Y et en effectuant des pseudo-divisions dans $K[X_1 \ldots X_n][Y]$. On remplace P_1 par P_1' et par le reste des pseudo-divisions de P_1 par $P_1', P_2, \ldots P_s$. On obtient donc la suite $P_1', P_2 \ldots P_s, R_1, R_2, \ldots, R_s$ avec

R_1 = pseudo-reste de P_1 par P_1'

R_i = pseudo-reste de P_1 par P_i pour $i = 2 \ldots s$.

Etant donné un tableau de signe de $P_1, P_2 \ldots P_s$ $R_1, \ldots R_s$ on en déduit celui de $P_1 \ldots P_s$ en utilisant uniquement les propriétés des polynômes sur un corps réels clos. Les principes utilisés sont les suivants :

(1) le signe de P_1' donne les variations de P_1

(2) le signe de R_1 donne le signe de P_1 au zéro de P_1'

(3) le signe de R_i $(i>1)$ donne le signe de P_1 au zéro de P_i

(4) le signe de P_1 à $-\infty$ est l'opposé du signe de P_1' à $(-\infty)$

(5) le signe de P_1 à $+\infty$ est le signe de P_1' à $+\infty$.

Et les zéros s'intercalent par le théorème des valeurs intermédiaires.

Pour comprendre comment on travaille sur les tableaux de signe, montrons tout de suite un exemple avec les notations introduites ensuite dans le problème de Kahan.

Soient P un polynôme de degré 1 en Y sur R[X] et E de degré 2 en Y sur R[X]. Supposons que pour $X=x$ fixé on ait le tableau de signe suivant pour la suite déduite de E et P

	y_0			y_1	
$\frac{\partial E}{\partial y}$	-	0	+	+	+
P	+	+	+	0	-
E mod $\frac{\partial E}{\partial Y}$	-	-	-	-	-
E mod P	+	+	+	+	-

On en déduit un unique tableau de signe pour E et P

		y_2		y_3		y_1	
E	+	0	-	0	+	+	+
P	+	+	+	+	+	0	-

y_0 n'apparait plus dans le tableau de signe de E et P, mais seulement les zéros de E et P : y_1, y_2, y_3. La nouvelle suite de polynômes a un polynôme de de moins de degré maximum en Y, donc en recommençant ce procédé on aboutira a une suite de polynômes de degré 0 en Y polynômes en $X_1 \ldots X_n$ que l'on appellera $B_1 \ldots B_m$.

Une fois obtenue la liste des polynômes $B_1 \ldots B_m$ en $\underline{X}$, il est possible en fixant les signes de ces polynômes, de reconstituer le tableau de signe des polynômes initiaux $P_1 \ldots P_s$ en $\underline{Y}$. La condition cherchée en $\underline{x}$ est alors la disjonction de toutes les conditions de signe sur les B_j telles que le tableau de signe des P_i qu'on en déduit donne une réponse positive à la question initiale (*).

Il faut ensuite traiter par la même méthode les cas où le coefficient dominant de certains polynômes est nul ce qui donne éventuellement un autre polynôme de degré maximum en Y et en tout cas d'autres calculs pour les pseudo-restes correspondant.

C'est long car le nombre des polynômes qui interviennent croît trop vite. Mais le passage d'une étape à l'autre est très simple.

I_2 - Algorithme de Seidenberg.

Seidenberg [S], commence par transformer les inégalités en une égalité $f = 0$, en ajoutant des variables ($P > 0$ se traduit par $\exists z \; Pz^2 = 1$), en faisant des sommes de carrés ($P_1 = 0$ et $P_2 = 0$ se traduisent par $P_1^2 + P_2^2 = 0$) ou des pro-

duits ($P_1 = 0$ ou $P_2 = 0$ se traduit par $P_1 P_2 = 0$). Puis il élimine les variables une par une par récurrence. Comme l'algorithme varie selon que le terme de plus haut degré de $f(X_1 \ldots X_n, X, Y)$ en Y : g est nul ou non, la récurrence est fondée sur un théorème qui montre l'existence de F et G tels que l'on ait l'équivalence entre

$$(\exists x)(\exists y) \text{ tel que } f(x_1 \ldots x_n, x, y) = 0 \text{ et } g(x_1 \ldots x_n, x) \neq 0$$

et

$$(\exists x) \text{ tel que } F(x_1 \ldots x_n, x) = 0 \text{ et } G(x_1 \ldots x_n) \neq 0 .$$

Une astuce de calcul transforme le formule f=0 et g≠0 en une seule égalité, et en fait, tout problème se ramène à remplacer l'existence d'un point réel sur une courbe par l'existence d'un x réel annulant un polynôme en X. Pour ceci, l'idée est de se limiter à l'existence d'un point réel parmi un nombre fini de points de la courbe, les points à distance extrêmale d'un point (a,o). Ceci nous donne une deuxième équation $f_a(x,y) = 0$. En choisissant a pour que le résultant de f et f_a en y ne soit pas identiquement nul (choix fini borné par le degré en Y), on obtient bien un nombre fini de points communs à f et f_a. Pour chaque zéro réel du résultant, on a au moins un zéro réel en y ou 2 zéros conjugués dans $K[\sqrt{-1}]$ en y pour f. On change la direction de l'axe des y pour éviter les zéros conjugués. L'élimination des pentes indésirables (racines d'un polynôme) se fait en plaçant à l'extérieur d'un intervalle borné, fonction des coefficients a_i du polynôme. Les coefficients de ce polynôme s'obtiennent par un calcul de fonctions symétriques de racines.

Alors l'existence d'un point réel (x,y) de f équivaut à l'existence d'un zéro réel x pour leur résultant.

La dernière variable ne peut s'éliminer de la même manière, on utilise cette fois le théorème de Sturm qui nous donne des inégalités. On ne pouvait

le faire dans les étapes précédentes sous peine de changer totalement la méthode de récurrence fondée sur = et ≠ et des calculs de résultants.

Pour le détail, on peut se référer à l'article de Seidenberg [S].

I_3 Algorithme de Collins.

L'élimination de n+1 variables se fait en trois étapes.

Etant donné $P_1 \ldots P_s \in \mathbb{Q}[X_1 \ldots X_n, Y]$, la première étape consiste d'abord à trouver des polynômes en $X_1 \ldots X_n$: $Q_1 \ldots Q_{p_1}$ tels que, au-dessus des semi-algébriques A_i de R^n déterminés pour ces polynômes en $X_1 \ldots X_n$, le nombre des zéros complexes des $P_1 \ldots P_s$ en Y soit constant, c'est-à-dire qu'un tableau donnant la répartition des zéros des P_i est le même pour tout $(x_1 \ldots x_n) \in A_i$. On sait alors que, sur chaque composante connexe de ce découpage, le tableau de signe des polynômes $P_1 \ldots P_s$ sera constant. Mais à cette étape on n'a pas le moyen, ni de distinguer les composantes connexes, ni de compter le nombre de zéros réels au-dessus de chaque composante connexe.

Ce calcul se fait encore par un procédé de pseudo-division successive (pseudo remainder sequence). Des calculs de PGCD déterminent les zéros communs à deux polynômes ou la multiplicité des zéros. Par récurrence, on continue cette première étape jusqu'à obtenir une suite de polynômes en X_1 à coefficients rationnels, mais sans savoir expliciter les conditions de signe à chaque étape.

La deuxième étape consiste en l'élimination de X_1 par un calcul "explicite" des zéros réels des polynômes en X_1 obtenus. Le théorème de Sturm permet d'en déterminer le nombre. Chaque zéro est alors un nombre algébrique réel se calculant pour son polynôme minimal sur $\mathbb{Q}$ et un intervalle d'extrê-

mité rationnelle séparant les racines. On connait alors le signe des polynômes en X_1 par un test en chaque racine algébrique réelle ou par un test en un point de chaque intervalle entre deux racines.

La troisième étape est la remontée. On détermine le signe des polynômes en X_1, X_2 pour chaque valeur de $X_1 : x_1$. Si x_1 n'est pas rationnel, on travaille sur $\mathbb{Q}(x_1)$. Pour les valeurs x_2 algébriques réelles non rationnelles, on doit alors calculer dans $\mathbb{Q}(x_1, x_2)$ qui, par le théorème de l'élément primitif, s'identifie à une extension $\mathbb{Q}(\alpha$. Les signes des polynômes en (x_1, x_2) se calculent alors comme dans la première étape. De cette manière, on peut continuer la remontée par récurrence jusqu'à X_n et Y. On obtient ainsi la décomposition algébrique cylindrique. On connait le signe des $P_1 \ldots P_s$ sur chaque morceau de R^{n+1}, donc on peut tester l'existence d'un $(x_1 \ldots x_n, y)$ tel que le signe des $P_1 \ldots P_s$ soit fixé.

Cette méthode semble au point lorsque toutes les variables à éliminer sont quantifiées. Par exemple, pour deux courbes à coefficients réels donnés, on peut déterminer le découpage du plan et donc les signes des 2 polynômes. Mais dans le problème de Kahar, le fait d'avoir à éliminer x_o, y_o, a, b qui ne sont pas quantifiées, rend l'algorithme difficilement praticable. Je n'ai donc pas cherché à mettre en oeuvre cet algorithme, d'autant plus qu'Arnon n'avait pas abouti dans le cas général.

La bibliographie sur cet algorithme est importante avec les articles de Arnon, Brown, Collins, Traub, Loos, MC Callum [ACM],[A],[BT],[C],[CL], [LO1],[LO2].

I_4 - Comparaison générale des algorithmes

La conception de Hormander est totalement différente car sa récurrence porte entièrement sur des inégalités et des résultats purement réels : théorème des valeurs intermédiaires, variation d'un polynôme en fonction du signe de sa dérivée, signe d'un polynôme à l'infini.

Par contre, Seidenberg et Collins utilisent dans leur récurrence des conditions ($= 0$) ou ($\neq 0$) et des arguments portant sur la clôture algébrique du corps réel clos : propriétés des résultants et des sous résultants. Ces calculs leur permettent de déterminer le nombre de zéros dans la clôture algébrique, les méthodes diffèrent ensuite pour distinguer les zéros conjugués dans la clôture algébrique, des zéros réels. Tous deux n'introduisent les inégalités que pour éliminer la dernière variable en appliquant le théorème de Sturm.

Toutefois, le résultat final obtenu par Collins et Hormander est analogue et donne un résultat plus complet que la question initiale puisqu'il détermine entièrement le signe des polynômes P_i dans le sens suivant.

Les deux méthodes aboutissent à un découpage de l'espace des $\underline{x}$: R^n

- en semi-algébriques définis par des conditions de signes sur les polynômes $B_1(\underline{X})...B_m(\underline{X})$ pour Hormander, partition non nécessairement connexe ;

- en semi-algébriques connexes homéomorphes à un pavé de R^n définis par la "remontée" des points et des intervalles de R pour Collins.

Le point commun à ces deux partitions c'est qu'au-dessus de chaque morceau semi-algébrique A_i de ce découpage, le tableau de signes en y des $P_i(\underline{x},y)$ est indépendant du $\underline{x}$ choisi dans A_i. Voyons sur un exemple très simple en quoi diffère le découpage.

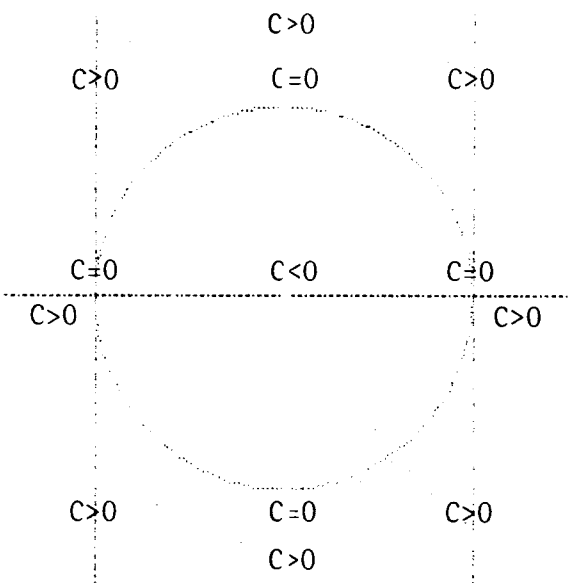
Pour trouver une condition sur x telle que

$$\exists y \quad x^2+y^2-1<0 .$$

Hormander aboutit en 2 étapes à la partition de R en $x^2-1<0$, $x^2-1=0$, $x^2-1>0$, soit 3 semi-algébriques. Puis, sans étudier le polynôme x^2-1 comme polynôme en x, il constate en regardant les tableaux de signes de x^2+y^2-1 comme polynôme en y pour chaque signe de x^2-1, s'il existe ou non un y tel que $x^2+y^2-1<0$.

Collins découpera R selon les racines de x^2-1 en points et intervalles $x<-1$, $x=-1$, $-1<x<1$, $x=1$, $x>1$. Soient 5 semi-algébriques connexes. La suite du travail consiste alors à connaître le nombre de racines en y et à tester le signe de x^2+y^2-1 au-dessus de chaque point ou intervalle dans chaque morceau obtenu. Ce travail est sensiblement identique que pour Hormander, mais il y a 5 tests à faire au lieu de 3.

C'est ce découpage que M. Coste a baptisé d'une manière très imagée "saucissonnage" dans [COS] et que Collins désigne par "cylinder algebraic decomposition" [C].



II - Le probleme de l'ellipse et du cercle

J'ai trouvé intéressant de tester les différents algorithmes sur ce problème concret pour en mieux voir les avantages et les inconvénients : nombre de calcul pour un algorithme général, nombre de tests maximum pour x_0, y_0, a, b donné quelconque, rapport entre les inégalités obtenues et les solutions géométriques que l'on a envie d'écrire.

II_1 - Hormander et le problème de Kahan

J'ai abordé le problème de Kahar en appliquart l'algorithme mais en éliminant des étapes inutiles.

Le problème étant écrit sous la forme naturelle

$$(\forall x)(\forall y)\ (E(x,y) = 0 \implies C(x,y) < 0)$$

où C et E sont des polynômes en $x, y, x_0, y_0, \frac{1}{a}$ et $\frac{1}{b}$ avec a et b non nuls. On s'intéresse au signe de C quand E est nul, donc la première démarche pour éliminer y est de faire la pseudo-division de C par E comme polynôme en y

$$\frac{1}{b^2} C = E + P$$

où P est une parable $P(x,y) = \frac{2y_0}{b^2}(y-y_c)+F(x)$.

On a alors à étudier le signe de P quand E est nul, ce qui nous conduit à écrire les suites de polynômes suivantes lorsqu'au cours des pseudo-divisions, aucun coefficient dominant ne s'annule.

Etape	Suites de polynômes	degré a priori en y	coefficient dominant en y
1	E	2	$1/b^2 \neq 0$
	P	1	$2y_o/b^2$
2	$\frac{\partial E}{\partial y} = \frac{2(y-y_o)}{b^2}$	1	$2/b^2$
	P	1	$2y_o/b^2$
	$I = E \bmod_y \frac{\partial E}{\partial y} = \frac{(x-x_o)^2}{a^2} - 1$	0	
	$R = (4y_o^2 / b^2\ E) \bmod_y F$	0	
3	$\frac{\partial P}{\partial y} = 2y_o/b^2$	0	
	$\frac{\partial E}{\partial y} = \frac{2(y-y_o)}{b^2}$	1	$1/b^2 \neq 0$
	$I = \frac{(x-x_o)^2}{a^2} - 1$	0	
	R		
	$F = [P \bmod_y \frac{\partial E}{\partial y}]$	0	
4	$\frac{\partial^2 E}{\partial y^2} = 2/b^2$	0	
	$\frac{\partial P}{\partial y} = 2y_o/b^2$	0	
	I	0	
	R	0	
	F	0	

avec $F(x) = \frac{x^2+y_o^2-1}{b^2} - [\frac{(x-x_o)^2}{a^2} - 1]$

et $R(x) = \frac{4y_o^2}{b^2}[\frac{(x-x_o)^2}{a^2} - 1] + F(x)^2$ = résultant y(E,C).

On remarque que F est de degré 2 en x et R de degré 4.

Dans le cas général $y_o \neq 0$, on aboutit à la condition suivante où y est éliminé

$$(\forall x) \qquad I(x) = \frac{(x-x_o)^2}{a^2} - 1 < 0 \Longrightarrow F(x) < 0 \quad \text{et} \quad R(x) > 0$$

par les arguments suivants qui respectent Hormander à la lettre.

Le signe de $\frac{\partial^2 E}{\partial y}$ montre que E passe par un minimum. En ce minimum, E est du signe de $I = E \bmod_y \frac{\partial E}{\partial y}$ et P du signe de $F = P \bmod_y \frac{\partial E}{\partial y}$. On veut obtenir P négatif dès que E s'annule. D'abord, pour que E s'annule, il faut et il suffit que son minimum I soit négatif. Pour que P soit négatif dès que E s'annule, comme P est monotone en y, il faut et il suffit que P soit négatif entre les deux racines de E en y (distinctes ou confondues), et pour cela, il faut et il suffit que

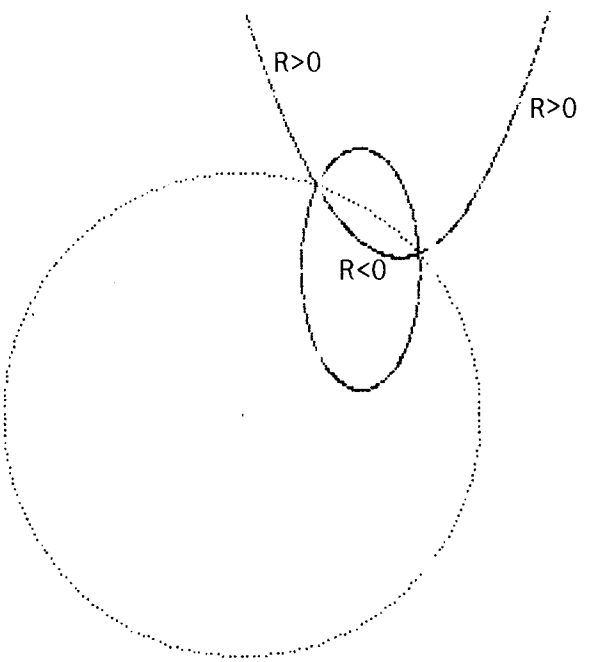
1) P soit négatif en un point entre les racines de E en y, par exemple au minimum de E, et

2) P s'annule à l'extérieur des racines de E en y, c'est-à-dire pour E>0.

Hormander remplace ces deux conditions par $F(x) < 0$ et $R(x) > 0$, ce qui aboutit à la condition sur x donnée.

Avant d'éliminer x, on peut regarder l'intérêt géométrique des conditions obtenues. Il est clair que pour résoudre la question, il suffit d'étudier x tel que $I(x) < 0$, le résultant R(x) est surement indispensable

pour connaître la position respective du cercle et de l'ellipse. Avec les divisions de l'algorithme, on peut voir en prenant un point x,y sur la parabole P(x,y) = 0, que le signe de R(x) est positif si (x,y) est extérieur au cercle et à l'ellipse, et négatif s'il est intérieur au deux, la parabole étant la seule courbe du faisceau de degré 1 en y. La condition $R(x) > 0$ pour $I(x) \leqslant 0$ équivaut en fait à ce que l'ellipse ne rencontre pas le cercle ($R(x) \neq 0$ pour $I(x) \leqslant 0$), la position (intérieure ou extérieure) va être donnée par le signe de F(x).



$x_o = 5$

$y_o = 6$

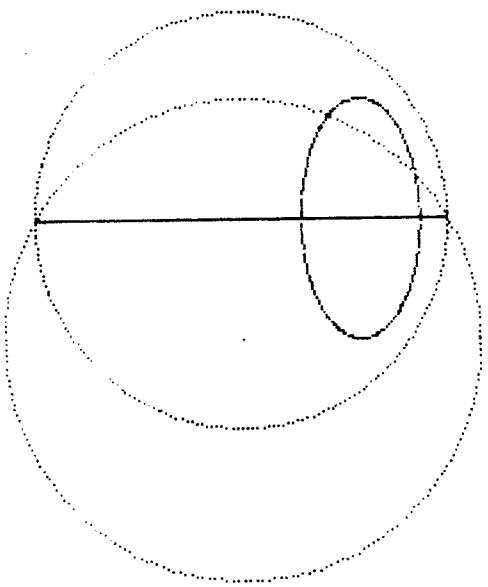
$a = 2,5$

$b = 5$

Etudions F(x), si l'on prend (x y) point de l'ellipse au-dessus de x, on voit que

$$x^2 + (y-y_o)^2 = 1 - y_o^2 + b^2F(x)$$

le signe de F(x) donne donc la position de l'ellipse par rapport à un cercle de centre $(0,y_o)$ et de rayon $1-y_o^2$, c'est le cercle coupant le cercle unité pour $y = y_o$. Si $F(x) < 0$, l'ellipse est intérieure à ce cercle, et pour $a < b$, elle ne peut couper le cercle unité qu'en 2 points. Cette remarque permettra de simplifier l'étude de la condition $R(x) > 0$ sachant que R a au plus 2 racines réelles.



$x_o = 0,5$

$y_o = 0,5$

$a = 0,25$

$b = 0,5$

On voit aussi que tout algorithme éliminant les variables l'une après l'autre fera jouer un rôle très dissymétrique aux variables, même dans un problème symétrique au départ.

L'élimination de x dans l'assertion

$$(\forall x)\ \left(\frac{(x-x_o)^2}{a^2} < 1 \Longrightarrow F(x) < 0 \quad \text{et} \quad R(x) > 0\right)$$

est un peu plus laborieuse.

Je l'ai menée jusqu'au bout. C'est d'ailleurs en faisant ce travail que j'ai vu l'intérêt de choisir d'éliminer y d'abord pour $a < b$. En effet, dans ce cas, j'ai remarqué que si $F < 0$ sur $I = [x_o-a, x_o+a]$ alors $R'' > 0$ sur I car

$$(R'' = \frac{4y_o^2}{a^2b^2} + 2(F'^2+FF'') \quad \text{et} \quad F'' = \frac{1}{b^2} - \frac{1}{a^2} < 0)$$

Le nombre de racines de R sur I est donc inférieur ou égal à 2. Pour $a > b$, on choisira d'éliminer x d'abord. On obtiendra les conditions en échangeant les rôles de x_o et y_o, a et b dans les conditions trouvées pour $a < b$.

Compte tenu de cette remarque j'étudie d'abord, pour $a < b$, le signe de F sur l'intervalle I. J'obtiens par l'algorithme des conditions portant sur $F'' < 0$, sur la position du minimum de F par rapport à I donné par $I \bmod F'$, et sur (le signe de F en x_o-a et x_o+a donné par $F \bmod I$ ou le signe de F en son minimum donné par le discriminant de F : $F \bmod F'$). Pour $a < b$, ceci donne deux conditions à tester pour avoir $F < 0$ sur I.

J'applique alors l'algorithme à R en ne continuant pas à dériver R" puisqu'il n'est jamais nul sur I. En 5 étapes de divisions successives, on obtient 5 polynômes en x_o, y_o, a et b dont le discriminant de R et les 2 coefficients dominants qui apparaissent lors des divisions successives de R par R' pour le calcul du discriminant. A ces 5 polynômes, il faut ajouter quelques conditions en x_o-a et x_o+a pour situer tout cela sur I. On obtient 10 nouvelles conditions au maximum pour le signe de R sur I.

Je peux donner ici, sans le détail des étapes successives, les résultats obtenus pour $y_o = 0$ qui montrent la finesse des conditions obtenues par l'algorithme

$$(a = b \quad \text{et} \quad x_o > 0 \quad \text{et} \quad x_o + a < 1)$$

$$\text{ou} \quad (a = b \quad \text{et} \quad x_o < 0 \quad \text{et} \quad x_o - a > -1)$$

$$\text{ou} \quad (a = b \quad \text{et} \quad x_o = 0 \quad \text{et} \quad b^2 < 1)$$

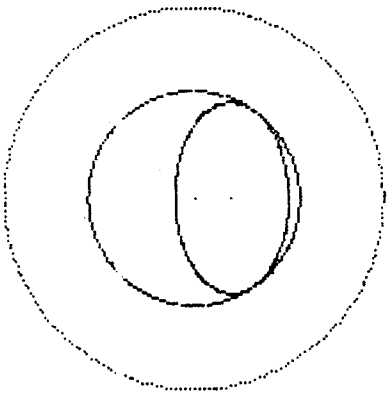
$$\text{ou} \quad (a < b \quad \text{et} \quad \frac{x_o a}{b^2 - a^2} \in [-1,+1] \quad \text{et discriminant} \quad (E \bmod C) < 0)$$

$$\text{ou} \quad (a < b \quad \text{et} \quad \frac{x_o a}{b^2 - a^2} \notin [-1,+1] \quad \text{et} \begin{cases} x_o \geqslant 0 \quad \text{et} \quad x_o + a < 1 \\ \text{ou} \\ x_o < 0 \quad \text{et} \quad x_o - a > -1 \end{cases}$$

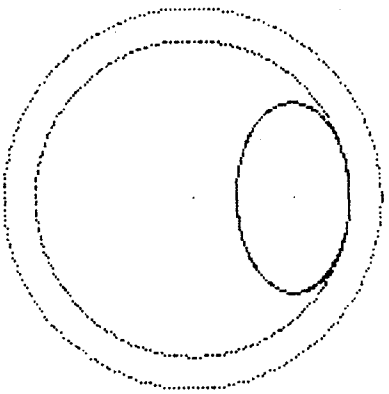
$$\text{ou} \quad a > b \quad \text{et} \begin{cases} x_o \geqslant 0 \quad \text{et} \quad x_o + a < 1 \\ \text{ou} \\ x_o < 0 \quad \text{et} \quad x_o - a > 1 \end{cases}$$

On voit clairement sur ces résultats l'importance du sens de l'ellipse horizontale ou verticale et la simplicité dans le cas horizontal.

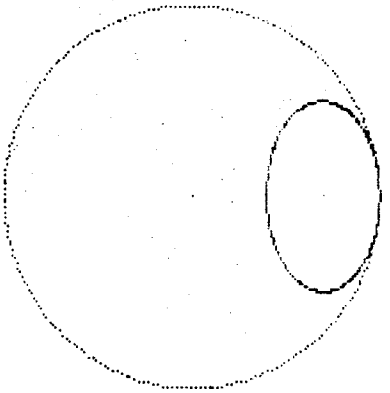
Dans le cas vertical, la condition plus compliquée pour $\frac{x_o a}{b^2 - a^2} \in [-1,+1]$ provient du fait que dans ce cas l'ellipse admet deux points à distance maximale de l'origine, alors que dans le cas contraire $\frac{x_o a}{b^2 - a^2} \notin [-1,+1]$ le point à distance maximale de l'origine est $(x_o - a)$ ou $(x_o + a)$



$x_o = 0,2$
$y_o = 0$
$a = 0,3$
$b = 0,5$
$\frac{x_o a}{b^2-a^2} = 0,375$



$x_o = 0,533$
$y_o = 0$
$a = 0,3$
$b = 0,5$
$\frac{x_o a}{b^2-a^2} = 1$



$x_o = 0,7$
$y_o = 0$
$a = 0,3$
$b = 0,5$
$\frac{x_o a}{b^2-a^2} = 1,312$

Au total, si l'on représente les tests à faire sous forme d'arbre, on obtient au maximum des branches de longueur 10 tests. Il y a beaucoup de ramifications, mais certaines branches sont très courtes et ne nécessitent pas les calculs de R et de son discriminant. Mais simplement quelques tests en (x_o-a) et (x_o+a). Le détail des résultats figurera dans ma thèse.

II_2 - Seidenberg et le problème de Kahan

Ma première démarche pour suivre l'idée de Seidenberg a été d'ajouter une variable pour supprimer l'inégalité. Mais il est devenu vite évident que c'était une très mauvaise méthode, vu la difficulté d'éliminer une variable supplémentaire. L'idée de prendre un point à distance extrêmale de l'origine sur l'ellipse n'a pas abouti non plus de manière satisfaisante, du fait que

je tenais à rester dans les principes d'un algorithme général. J'ai finalement abouti à la méthode suivante de conception en apparence très simple.

L'idée est d'utiliser le résultant du cercle et de l'ellipse pour écrire que l'ellipse ne rencontre pas le cercle en ajoutant qu'un point de l'ellipse particulier est intérieur au cercle par exemple $(x_o+a,0)$ (condition ne comportant plus de variables quantifiées).

Pour que l'ellipse et le cercle n'aient pas de zéro commun dans $\mathbb{C}$ en y, il faut et il suffit que le résultant soit non nul. Mais ici il est facile de distinguer pour x_1 fixé réel si les zéros en y sont réels ou complexes conjugués, car si x_1 est zéro réel du résultant, $y_1^2 = 1-x_1^2$ donne des points réels communs au cercle et à l'ellipse si et seulement si $x_1 \in [-1,+1]$. La question initiale est alors équivalente à la condition sur (x_o+a) : $(x_o+a)^2-1 < 0$, et R(x) n'a pas de zéro réel sur [-1,+1] : ce qui se calcule aisément par le théorème de Sturm.

On aboutit, pour le calcul de la suite standard de R, aux mêmes divisions successives au signe près que dans Hormander. En travaillant un peu, on peut obtenir dans le cas général 10 conditions à tester un peu moins que pour Hormander. Mais ici le cas $y_o = 0$ par exemple rentre dans le calcul commun (si l'on calcule le résultant comme un déterminant). Si l'on regarde les conditions disposées dans un arbre, le nombre de branches est petit, mais presque toutes les branches sont de longueur 10.

Pour conclure sur ces essais s'il semble d'abord que la méthode de Seidenberg soit beaucoup plus simple, en fait Hormander en considérant tous les cas possibles, permet dans beaucoup de cas d'obtenir un très petit nombre de tests, tests variés adaptés à chaque cas de figure. Quand on voit l'énormité du dernier reste de la suite de Sturm qui n'est autre que le discriminant du résultant, on est convaincu de l'utilité d'en éviter le calcul quand cela n'est pas indispensable. A titre indicatif, la dernière division pour obtenir ce reste prend 200000 ms de calcul pour Macsyma.

Je n'ai pas cherché à évaluer de manière précise la complexité des différents algorithmes. Il apparait sur l'exemple de Kahan que la complexité de l'algorithme de Hormander peut être réduite de plusieurs manières. D'abord, en commençant systématiquement par des divisions pour les polynômes soumis à la condition $P = 0$, on fait descendre le degré de l'ensemble des polynômes. Je ne vois pas de méthode permettant de réduire le nombre de polynômes considérés. Par contre, il faudrait établir des règles générales pour partir des conditions de signes que l'on veut obtenir sur les polynômes $P_1 \ldots P_s$ en Y pour, à chaque étape, ne considérer que les tableaux de signes pouvant aboutir à ces conditions. Sinon le nombre de cas à considérer est très grand, et de nombreux tableaux de signes sur les polynômes en $\underline{X}$ peuvent s'avérer impossible compte tenu des relations liant ces "constantes" en Y.

Quant à Collins, il doit être performant si l'on a une ellipse donnée à tester avec des valeurs numériques. Mais cet algorithme répond mal à l'élimination des quantificateurs du point de vue général, en conservant les variables non quantifiées.

Bibliographie

[ACM] D.S. ARNON, G.E. COLLINS and S. MC CALLUM : Cylindrical algebraic decomposition I et II : the basic algorithm, Siam J. Comput., Vol. 13 n° 4, nov. 84, p. 865-889.

[A] D.S. ARNON : Towards Mechanical Solution of Kahan ellipse problème I, Computer Algebra, Lectures Notes 162, Springer Verlag 1983.

[BCR] J. BOCHNAK, M. COSTE, M.-F. ROY : A paraître.

[BT] W.S. BROWN, J.-F. TRAUB : On Euclid's algorithm and the theory of subresultants, J. Assoc. Comput. Math. 18, 4 (1971), p. 505-514.

[C] G.E. COLLINS : Quantifier elimination for real closed fields a guide to the literature, Computer Algebra Symbolic and Algebraic Computation, Springer Verlag (1982-1983).

[CL] G.E. COLLINS, R. LOOS : Real zeros of polynomials, Computer Algebra Symbolic and Algebraic Computation, Springer Verlag (1982-1983).

[COS] M. COSTE : Ensembles semi-algébriques, Géométrie Algébrique Réelle et Formes quadratiques, Lecture Notes n° 959, Springer Verlag.

[H] HORMANDER : The analysis of linear partiel differential operators, tome 2, Springer Verlag (1983).

[J] JACOBSON : Basic Algebra I.

[Ka] W. KAHAN : "Problem # 9 : an ellipse problem", SIGSAM Bulletin of the assoc. Comp. Math. 9, p. 11 (1975).

[Kr] G. KREISEL and J.-L. KRIVINE : Elements of mathematical logic (model theory) North. Holland Amsterdam (1967).

[Lau] M. LAUER : "A solution to Kahan's problem (SIGSAM problem n° 9) "SIGSAM Bulletin of the Ass. Com. Math. 11, p. 16-20 (1977).

[Laz] D. LAZARD : Solution au problème de Kahan (non publié).

[Lo_1] R. LOOS : Generalized polynomial remander sequences, Computer Algebra Symbolic and Algebraic Computation, Springer Verlag (1982-1983).

[Lo_2] R. LOOS : Computing in algebraic extensions, Computer Algebra Symbolic and Algebraic Computation, Springer Verlag (1982-1983).

[M] M. MIGNOTTE : Solution au problème de Kahan (non publié).

[S] A. SEIDENBERG : A new decision method for elementary algebra Ann of Math. 60, 365-374 (1954).

Annette PAUGAM
Université de RENNES I
Département de Mathématiques
Campus de Beaulieu
35042 - RENNES CEDEX (**France**)