Fabio BERTO

Systèmes non-linéaires et optimisation

(notes du cours)

première édition
Genève, MIM – MM

Introduction

Ce document est basé sur mes notes du cours à option de deuxième cycle "Systèmes non linéaires et optimisation", donné par le Professeur Ernst Hairer (section de mathématiques de l'université de Genève) durant l'année académique 1999-2000.

Malheureusement, étant donné que, lorsque j'ai commencé à écrire ce polycopié, mon but principal n'était pas la rédaction d'un cours complet mais l'apprentissage de LaTeX et vu que, plus tard, le temps à ma disposition a diminué à cause de mes nombreux engagements, le lecteur ne trouvera ici que le premier chapitre traité au cours, ainsi qu'une bibliographie du sujet.

D'autre part, j'aimerai remercier Ernst Hairer, pour la qualité son cours. Je le remercie aussi pour la patience, dont il a fait preuve, en relisant intégralement mon travail et en me proposant plusieurs améliorations. Je remercie aussi Isabel Mendez et Claude Pache, qui ont supervisé l'orthographe et la grammaire, ainsi que Pierre Leone, qui a géré les séances d'exercices, où j'ai pu trouver plusieurs exemples et résultats intéressants.

Fabio Berto & Linh Pham [1]

1. Linh Pham a été un parmi ceux, qui m'ont pratiquement récrit l'introduction. Il m'a aussi proposé de rédiger ma biographie, mais pour l'instant je n'ai pas le temps.

Table des matières

Chapitre 1

Systèmes polynomiaux à plusieurs variables

Le but de ce chapitre consiste dans l'étude de systèmes polynomiaux à plusieurs variables sur un corps. Souvent nous écrirons un système

$$(\mathcal{F})\begin{cases} f_1(x_1, x_2, \ldots, x_n) = 0 \\ \vdots \\ f_m(x_1, x_2, \ldots, x_n) = 0 \end{cases} \tag{1.1}$$

où les f_i sont des polynômes à n variables et à coefficients dans un corps commutatif $\mathbb{K}$, i.e.

$$f_i \in \mathbb{K}[x_1, \ldots, x_n] \ \ \forall\ i = 1, \ldots, m$$

sous la forme

$$\{f_i(x_1, \ldots, x_n) = 0\}_{i=1}^m$$

En l'occurrence $\mathbb{K}$ joue le rôle de $\mathbb{R}$, $\mathbb{Q}$, ou $\mathbb{C}$.

Si $\mathbb{K} = \mathbb{C}$ il y a une jolie propriété. Nous pouvons toujours factoriser les polynômes à une variable. Par contre si $\mathbb{K} = \mathbb{Q}$, alors nous pouvons supposer que les f_i soient à coefficients entiers, tout simplement en les multipliant par le ppcm des dénominateurs des coefficients. Donc nous pouvons éviter tout problème d'arrondis!

Il y a principalement deux questions, auxquelles nous pouvons essayer de répondre. La première serait de savoir si, étant donné un système d'équations polynomiales, il possède une solution. La deuxième est de déterminer, dans ce cas, s'il est possible de trouver toutes les solutions, ou au moins une partie de celles-ci.

1.1 La matrice de Sylvester

Regardons ces deux polynômes à une variable et à coefficients dans un corps $\mathbb{K}$

$$\begin{aligned} f(x) &= a_0 + a_1 x + \cdots + a_k x^k \\ g(x) &= b_0 + b_1 x + \cdots + b_\ell x^\ell \end{aligned}$$

avec a_k et b_ℓ non nuls. La question qu'on se pose c'est de savoir si le système

$$\begin{cases} f(x) = 0 \\ g(x) = 0 \end{cases}$$

admet des solutions. Une réponse immédiate nous dit que le système admet une solution si et seulement si $\deg[\mathrm{pgcd}(f(x), g(x))] \geq 1$. En effet, si une solution $\hat{x}$ existe, alors $\hat{x}$ serait une racine commune de $f(x)$ et de $g(x)$. De plus, le polynôme $x - \hat{x}$ diviserait chaque équation dans le système.

Définition 1. *Pour deux polynômes $f(x)$ et $g(x)$ comme ci-dessus, définissons la* ***matrice de Sylvester*** *de la façon suivante:*

$$Syl(f,g) = \begin{array}{c} \left(\begin{array}{cccccc} a_0 & \dots & a_k & & & \\ & \ddots & & & \ddots & \\ & & a_0 & \dots & & a_k \\ b_0 & \dots & b_\ell & & & \\ & \ddots & & & \ddots & \\ & & b_0 & \dots & & b_\ell \end{array} \right) \end{array} \begin{array}{l} \left.\vphantom{\begin{array}{c} a \\ a \\ a \end{array}}\right\} \ell \\ \left.\vphantom{\begin{array}{c} a \\ a \\ a \end{array}}\right\} k \end{array}$$

Remarquons que la matrice de Sylvester est une matrice carrée. Nous pouvons donc calculer son déterminant. D'où la définition suivante.

Définition 2. *Le* ***résultant*** *de deux polynômes $f(x)$ et $g(x)$ noté par*

$$Res(f,g)$$

est le déterminant de la matrice de Sylvester $Syl(f,g)$.

Le théorème suivant nous donne l'équivalent de l'identité de Bézout [1] pour calculer le pgcd de deux polynômes à une variable. En fait, nous verrons plus tard qu'elle se généralise à plusieurs polynômes avec une quantité finie de variables.

Théorème 1. *Soient $f(x)$ et $g(x) \in \mathbb{K}[x]$, avec $\deg(f) = k$ et $\deg(g) = \ell$.*
Alors $\exists$ $p(x)$ et $q(x) \in \mathbb{K}[x]$, non nuls avec $\deg(p) < \ell$ et $\deg(q) < k$, tels que

$$f(x)p(x) + g(x)q(x) = Res(f,g)$$

1. Identité de Bézout: Soit $\{a_i\}_{i=1}^m$ une famille d'entiers. Alors, pour que ces entiers soient premiers entre-eux, il faut et il suffit qu'il existe d'autres entiers b_i, $i = 1, \dots, m$, tels que $\sum_{i=1}^m a_i b_i = 1$.

Démonstration. Soient $f(x)$ et $g(x) \in \mathbb{K}[x]$, $\deg(f) = k$ et $\deg(g) = \ell$. Observons le calcul matriciel suivant:

$$Syl(f,g) \cdot \begin{pmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^{k+\ell-1} \end{pmatrix} = \begin{pmatrix} f(x) \\ xf(x) \\ \vdots \\ x^{\ell-1}f(x) \\ g(x) \\ xg(x) \\ \vdots \\ x^{k-1}g(x) \end{pmatrix} \tag{1.2}$$

Nous avons deux cas:

1. $Res(f,g) \neq 0$: Dans ce cas $Syl(f,g)$ est inversible. Nous pouvons donc multiplier l'expression par $Syl(f,g)^{-1}$. La première ligne nous donne

   $$1 = \tilde{p}(x)f(x) + \tilde{q}(x)g(x)$$

   Celle-ci multipliée par $Res(f,g)$ livre le résultat cherché.

2. $Res(f,g) = 0$: Dans ce cas 0 est une valeur propre de $Syl(f,g)$. Il existe donc un vecteur propre ligne pour cette valeur propre, i.e.

   $$\underbrace{(p_0, \ldots, p_{\ell-1}, q_0, \ldots, q_{k-1})}_{\neq 0} \cdot Syl(f,g) = 0$$

   Si nous multiplions (1.2) par ce vecteur, la partie de droite nous offre un produit scalaire, qui se traduit comme suit:

   $$p(x)f(x) + q(x)g(x) = 0$$

   avec $p(x) = \sum_{i=0}^{\ell-1} p_i x^i$, et $q(x) = \sum_{i=0}^{k-1} q_i x^i$.

□

De cette démonstration nous pouvons tirer les deux conclusions suivantes.

1. Si $Syl(f,g)$ est inversible, alors $f(x)$ et $g(x)$ sont premiers entre eux. Donc ils n'ont pas de racines communes.
   Donc le système $\{f(x) = 0, g(x) = 0\}$ n'admet pas de solution.

2. Si $Syl(f,g)$ est singulière, nous pouvons trouver des polynômes $p(x)$ et $q(x)$, avec $\deg(p) < \deg(g)$ et $\deg(q) < \deg(f)$, tels que $p(x)f(x) + q(x)g(x) = 0$.
   Donc $p(x)f(x) = -q(x)g(x)$. Ce qui implique que chaque racine de $f(x)$ est aussi une racine de $-q(x)g(x)$. Mais comme $\deg(f) > \deg(q)$, il doit donc forcément exister une racine de $f(x)$, qui soit aussi une racine de $g(x)$, et vice versa.

Théorème 2. *Soient $f(x)$ et $g(x) \in \mathbb{K}[x]$. Soient $k = \deg(f)$ et $\ell = \deg(g)$. Alors*

$$\dim \ker Syl(f,g) = \deg \operatorname{pgcd}(f(x), g(x))$$

Autrement dit: le nombre de solutions du système $\{f(x) = 0, g(x) = 0\}$ est égal à $k + \ell - \operatorname{rang} Syl(f,g)$.

Démonstration. Démontrons les deux inégalités:

1. $\dim \ker Syl(f,g) \geq \deg \operatorname{pgcd}(f(x), g(x))$:
   Supposons que $m = \deg \operatorname{pgcd}(f(x), g(x))$ et que $z_1, \ldots, z_m$ soient les racines du $\operatorname{pgcd}(f(x), g(x))$. Alors $f(z_i) = g(z_i) = 0$, $\forall\, i = 1, \ldots, m$. Supposons, pour l'instant, que les z_i soient distincts. Donc

$$\begin{pmatrix} 1 \\ z_1 \\ \vdots \\ z_i^{k+\ell-1} \end{pmatrix} \in \ker Syl(f,g) \;\; \forall i = 1, \ldots, m$$

   D'où $\dim \ker Syl(f,g) \geq m$.
   D'autre part, s'il existe un z_i, racine multiple avec multiplicité μ_i, alors z_i est aussi racine de tous les polynômes dérivés $\operatorname{pgcd}(f(x), g(x))^{(\nu_i)}$, $\nu_i = 1, \ldots, \mu_i - 1$. Ceci implique qu'il est aussi racine de $f^{(\nu_i)}(x)$ et de $g^{(\nu_i)}(x)$, $\forall\, \nu_i = 1, \ldots, \mu_i - 1$.
   Dans ce cas, l'idée est de dériver l'identité (1.2). Nous obtenons

$$Syl(f,g) \cdot \begin{pmatrix} 0 \\ 1 \\ 2x \\ \vdots \\ (k+\ell-1)x^{k+\ell-2} \end{pmatrix} = \begin{pmatrix} f'(x) \\ xf'(x) + f(x) \\ \vdots \\ x^{\ell-1}f'(x) + (\ell-1)x^{\ell-2}f(x) \\ g'(x) \\ xg'(x) + g(x) \\ \vdots \\ x^{k-1}g'(x) + (k-1)x^{k-2}g(x) \end{pmatrix}$$

   D'où, si nous remplaçons x par z_i,

$$\begin{pmatrix} 0 \\ 1 \\ 2z_i \\ \vdots \\ (k+\ell-1)z_i^{k+\ell-2} \end{pmatrix} \in \ker Syl(f,g)$$

   Ce dernier vecteur colonne est linéairement indépendent des autres[2].

2. Cette affirmation est facilement démontrable par récurrence en utilisant le déterminant d'une matrice de Vandermonde, dans laquelle nous avons dérivé une colonne.

2. $\dim \ker Syl(f,g) \leq \deg \operatorname{pgcd}(f(x), g(x))$:
   Soit $m = \dim \ker Syl(f,g)$. Si $m = 0$, alors $Syl(f,g)$ est inversible, donc, par le théorème 1, $\operatorname{pgcd}(f(x), g(x))$ est une constante non nulle. Si $m \neq 0$, il existe m vecteurs lignes $v_1^T, \ldots, v_m^T$ linéairement indépendents tels que $v_i^T Syl(f,g) = 0$, $i = 1, \ldots, m$. Ceci donne l'identité matricielle suivante:

$$\begin{pmatrix} v_1^T \\ \vdots \\ v_m^T \end{pmatrix} \cdot Syl(f,g) = 0$$

   Soit

$$\begin{aligned} v &= (p_0, \ldots, p_{\ell-1}, q_0, \ldots, \ldots, \ldots, q_{k-1}) \\ &= (p_0, \ldots, p_{\ell-1}, q_0, \ldots, q_{k-m}, \underbrace{0, \ldots, 0}_{m-1}) \end{aligned}$$

   une combinaison linéaire des vecteurs v_i. Comme précédement, si nous multiplions à gauche l'identité (1.2) par ce vecteur v, elle nous livre l'égalité

$$0 = \underbrace{p(x)}_{\deg \leq \ell-1} \underbrace{f(x)}_{\deg = k} + \underbrace{q(x)}_{\deg \leq k-m} \underbrace{g(x)}_{\deg = \ell}$$

   Factorisons $p(x)$, $f(x)$, $q(x)$ et $g(x)$ dans $\mathbb{C}[x]$. Comme $p(x)f(x) = -q(x)g(x)$, chacune des racines complexes de $f(x)$ doit forcément être une racine complexe de $q(x)g(x)$. Il y en a k, prises avec leurs multiplicités. Mais $\deg q(x) \leq k - m$. Donc il y a au moins m racines complexes de $f(x)$, prises avec leurs multiplicités, qui sont aussi racines de $g(x)$. Ce qui implique que $\deg \operatorname{pgcd}(f(x), g(x)) \geq m$.

$\square$

Corollaire 1. *Le système $\{f_i(x) = 0\}_{i=1}^m$ est sans solution si et seulement s'il existe des $p_i(x) \in \mathbb{K}[x]$ tels que $\sum_{i=1}^m f_i(x)p_i(x) = 1$.*

Démonstration. Pour la suffisance, la démonstration est triviale et se fait par la contraposée. Pour la nécessité, nous pouvons procéder par récurrence.
Pour $m = 2$ c'est le théorème 1. Supposons donc l'énoncé vrai pour tout système avec $m-1$ équations. Montrons-le pour $\{f_i(x) = 0\}_{i=1}^m$.
Soient $g_i(x) = \operatorname{pgcd}(f_i(x), f_m(x))$, $i = 1, \ldots, m-1$.
Or, chaque couple $\frac{f_i(x)}{g_i(x)}$ et $\frac{f_m(x)}{g_i(x)} \in \mathbb{K}[x]$ est sans racines communes, $\forall\ i = 1, \ldots, m-1$.
Donc $\exists\ \tilde{p}_i(x), \tilde{q}_i(x) \in \mathbb{K}[x]$ tels que

$$\frac{f_i(x)}{g_i(x)}\tilde{p}_i(x) + \frac{f_m(x)}{g_i(x)}\tilde{q}_i(x) = 1$$

Donc

$$f_i(x)\tilde{p}_i(x) + f_m(x)\tilde{q}_i(x) = g_i(x)$$

Or, le système $\{g_i(x) = 0\}_{i=1}^{m-1}$ n'a pas de solutions dans $\mathbb{K}$. Donc, par hypothèse de récurrence, $\exists\ r_i(x) \in \mathbb{K}[x]$ tels que

$$\sum_{i=1}^{m-1} r_i(x)g_i(x) = 1$$

D'où

$$\sum_{i=1}^{m-1} \underbrace{(r_i(x)\tilde{p}_i(x))}_{=:p_i(x)} f_i(x) \ + \underbrace{(\sum_{i=1}^{m-1} r_i(x)\tilde{q}_i(x))}_{=:p_m(x)} f_m(x) = 1$$

$\square$

1.2 Elimination de Kronecker et Nullstellensatz de Hilbert

"Like two beans in the bags!"
(Laurel & Hardy, *The Sons of the Desert*, 1933)

Le but de ce paragraphe consiste à généraliser le corollaire 1 pour des polynômes à plusieurs variables.
Le théorème suivant, dû à Kronecker (1882), nous convient. En vérité ce n'est pas vraiment le théorème en soi, qui est important, mais sa démonstration, laquelle nous livre un algorithme pour résoudre un système d'équations polynomiales à plusieurs variables.

De ce théorème il existe plusieurs preuves. Nous allons présenter ici la méthode de Kronecker, reprise par Buchberger (1965), laquelle a retrouvé son importance suite au développement des ordinateurs.

Théorème 3. ***(Elimination de Kronecker, 1882)***
Soit $\mathbb{K}$ un corps. Soient $f_i \in \mathbb{K}[x_1, \ldots, x_n]$, $i = 1, \ldots, m$. Alors le système polynomial $\{f_i(x_1, \ldots, x_n) = 0\}_{i=1}^{m}$ n'a pas de solution, si et seulement s'il existe des $p_i \in \mathbb{K}[x_1, \ldots, x_n]$ tels que $\sum_{i=1}^{m} p_i(x)f_i(x) = 1$, où $x = (x_1, \ldots, x_n)$.

Démonstration. La suffisance est évidente. En effet, si le système admettait une solution ξ, alors $\sum_{i=1}^{m} p_i(\xi)f_i(\xi) = 0\ \forall\ p_i \in \mathbb{K}[x_1, \ldots, x_n]$.
L'idée pour démontrer la nécessité consiste à éliminer les variables à une à une, en introduisant des nouvelles, puis à résoudre le système final à une variable. Ensuite, nous pouvons remonter vers la solution, ou reprendre dès le début avec d'autres éliminations.

En pratique, soit le système polynomial (1.1)

$$(\mathcal{F})\begin{cases} f_1(x_1, x_2, \ldots, x_n) = 0 \\ \qquad\vdots \\ f_m(x_1, x_2, \ldots, x_n) = 0 \end{cases}$$

Procédons par récurrence sur n. Si $n = 1$ c'est le corollaire 1. Passons à $n \geq 2$. Soient

$$a(x_n) = \sum_{i=1}^{m} u_i f_i(x_1, \ldots, x_n)$$

$$b(x_n) = \sum_{i=1}^{m} v_i f_i(x_1, \ldots, x_n)$$

avec $a \in \mathbb{K}[x_1, \ldots, x_{n-1}, u_1, \ldots, u_m][x_n]$ et $b \in \mathbb{K}[x_1, \ldots, x_{n-1}, v_1, \ldots, v_m][x_n]$. Ils ont une seule variable (c'est x_n) et leurs coefficients sont des polynômes sur $\mathbb{K}$ à $m + n - 1$ indéterminées.
Donc

$$a(x_n) = a_0(x_1, \ldots, x_{n-1}, u_1, \ldots, u_m) + \cdots + a_k(x_1, \ldots, x_{n-1}, u_1, \ldots, u_m) x_n^k$$
$$b(x_n) = b_0(x_1, \ldots, x_{n-1}, v_1, \ldots, v_m) + \cdots + b_k(x_1, \ldots, x_{n-1}, v_1, \ldots, v_m) x_n^k$$

$$\begin{aligned} Res(a,b) &= \det \begin{pmatrix} a_0(x_1, \ldots, x_{n-1}, u_1, \ldots, u_m) & a_1(x_1, \ldots, x_{n-1}, u_1, \ldots, u_m) & \cdots & \\ & a_0(x_1, \ldots, x_{n-1}, u_1, \ldots, u_m) & \ddots & \ddots \\ & \ddots \quad \ddots & \ddots & \ddots \\ b_0(x_1, \ldots, x_{n-1}, v_1, \ldots, v_m) & b_1(x_1, \ldots, x_{n-1}, v_1, \ldots, v_m) & \cdots & \\ & b_0(x_1, \ldots, x_{n-1}, v_1, \ldots, v_m) & \ddots & \ddots \\ & \ddots \quad \ddots & \ddots & \ddots \end{pmatrix} \\ &= \sum_{\substack{i_1, \ldots, i_m \\ j_1, \ldots, j_m}} (\prod_{t=1}^{m} u_t^{i_t} v_t^{j_t}) \cdot g_{i_1 \cdots i_m j_1 \cdots j_m}(x_1, \ldots, x_{n-1}) \end{aligned} \tag{1.3}$$

Ce qui donne le système

$$(\mathcal{G})\begin{cases} g_{i_1 \ldots i_m j_1 \ldots j_m}(x_1, \ldots, x_{n-1}) = 0 \\ \\ \quad i_1 \geq 0, \ldots, i_m \geq 0 \\ \quad j_1 \geq 0, \ldots, j_m \geq 0 \end{cases} \tag{1.4}$$

Remarque: Il est très probable que le système $(\mathcal{G})$ contient beaucoup plus d'équations que le système initial $(\mathcal{F})$. Par contre, nous pouvons y trouver les mêmes polynômes plusieurs fois à une constante multiplicative près. Le système $(\mathcal{G})$ se réduit donc à ℓ équations, sensiblement moins nombreuses qu'avant.

Affirmation: les systèmes $(\mathcal{F})$ et $(\mathcal{G})$ sont équivalents [3], i.e. si $(\xi_1, \ldots, \xi_n)$ est une solution de $(\mathcal{F})$, alors $(\xi_1, \ldots, \xi_{n-1})$ est une solution de $(\mathcal{G})$, et si $(\xi_1, \ldots, \xi_{n-1})$ est une solution de $(\mathcal{G})$, alors il existe un ξ_n tel que $(\xi_1, \ldots, \xi_n)$ soit une solution de $(\mathcal{F})$.

Démonstration de l'affirmation en deux étapes:

1. Voyons que si $(\xi_1, \ldots, \xi_n)$ est une solution de $(\mathcal{F})$, alors $(\xi_1, \ldots, \xi_{n-1})$ est une solution de $(\mathcal{G})$.
   Soit A une matrice carrée. Si A est inversible, alors
   $$A^{-1} = \frac{1}{\det(A)}\,\widehat{A}$$
   où $\widehat{A}$ est la matrice des mineurs de A. Donc, si $\vec{v}$ et $\vec{w}$ sont deux vecteurs colonnes tels que $A\vec{v} = \vec{w}$, cette écriture est équivalente à
   $$\det(A) \cdot \vec{v} = \widehat{A}\vec{w}$$
   De plus cette dernière reste vraie même si A n'est pas inversible.
   Appliquons-la à l'identité (1.2), où f a été remplacé par a et g par b. Nous obtenons
   $$Res(a,b) \cdot \begin{pmatrix} 1 \\ x_n \\ x_n^2 \\ \vdots \\ x_n^{2k-1} \end{pmatrix} = \widehat{Syl(a,b)} \cdot \begin{pmatrix} a(x_n) \\ x_n a(x_n) \\ \vdots \\ x_n^{k-1} a(x_n) \\ b(x_n) \\ x_n b(x_n) \\ \vdots \\ x_n^{k-1} b(x_n) \end{pmatrix} \tag{1.5}$$
   L'implication est évidente si nous regardons la première ligne de (1.5) et que nous la comparions avec (1.3). Comme les u_i et les v_i sont arbitraires, il faut bien que
   $$g_t(x_1, \ldots, x_{n-1}) = \sum_{i=1}^{m} \alpha_{it}(x_1, \ldots, x_n) f_i(x_1, \ldots, x_n)$$

2. Soit maintenant $(\xi_1, \ldots, \xi_{n-1})$ une solution de $(\mathcal{G})$.
   Si $f_i(\xi_1, \ldots, \xi_{n-1}, \cdot) \equiv 0$, $\forall\ i = 1, \ldots, m$, nous avons terminé. Sinon nous pouvons supposer que $f_1(\xi_1, \ldots, \xi_{n-1}, \cdot) \not\equiv 0$. Posons $v_1 = 1$ et $v_i = 0$, $i = 2, \ldots, m$ et fixons $x_i = \xi_i$, $i = 1, \ldots, n-1$. Ceci implique que
   $$a(x_n) = \sum_{i=1}^{m} u_i f_i(\xi_1, \ldots, \xi_{n-1}, x_n)$$

3. L'équivalence entre les deux systèmes est importante, car c'est grâce à elle que nous pouvons être sûrs que, quand nous passons d'un système à l'autre, nous ne perdons pas de solutions.

$$b(x_n) = f_1(\xi_1, \dots, \xi_{n-1}, x_n)$$

Nous savons par (1.3) que $Res(a, b) = 0$. Donc

$$\begin{cases} a(x_n) = 0 \\ b(x_n) = 0 \end{cases}$$

admet au moins une solution [4]. Soient $\xi_n^{(1)}, \dots, \xi_n^{(r)}$ toutes les solutions distinctes de ce système. Elles sont en quantité finie car $a(x_n)$ et $b(x_n)$ sont des polynômes non nuls.
Il s'agit maintenant de montrer qu'au moins une de ces solutions soit une solution commune des $f_i(\xi_1, \dots, \xi_{n-1}, x_n)$, $i = 1, \dots, m$.
Pour résoudre ce problème, fixons des η_i distinctes, $i = 1, \dots, m$.
Soient $u_i := \eta_i$, $\forall\ i = 1, \dots, m$. Il existe un $\xi_n^{(j_1)}$ qui est solution de ce système. Soient maintenant $u_i := \eta_i^2$, $\forall\ i = 1, \dots, m$. Il existe $\xi_n^{(j_2)}$ qui est solution de ce système, etc.
Nous pouvons continuer comme cela à l'infini. Mais les $\xi_n^{(j)}$ sont en quantité finie. Donc il y a forcément un $\xi_n^{(j)}$, qui apparaît m fois pour certaines puissances des η_i. Disons $\xi_n^{(1)}$.
Soit $\xi_n := \xi_n^{(1)}$. Nous avons par définition de $a(x_n)$:

$$\underbrace{\begin{pmatrix} \eta_1^{t_1} & \dots & \eta_m^{t_1} \\ \vdots & \star & \vdots \\ \eta_1^{t_m} & \dots & \eta_m^{t_m} \end{pmatrix}}_{=:A} \cdot \begin{pmatrix} f_1(\xi_1, \dots, \xi_n) \\ \vdots \\ f_m(\xi_1, \dots, \xi_n) \end{pmatrix} = 0$$

avec A inversible. Ceci implique que $f_i(\xi_1, \dots, \xi_n) = 0\ \forall\ i = 1, \dots, m$.

L'affirmation termine la preuve, car

$$\begin{array}{lcl} \text{($\mathcal{F}$) n'a pas de solutions} & \overset{aff.}{\Longleftrightarrow} & (\mathcal{G}) \text{ n'a pas de solutions} \\ & \overset{\substack{hyp.\\ réc.}}{\Longleftrightarrow} & \exists\ q_t \in \mathbb{K}[x_1, \dots, x_{n-1}] \text{ tels que} \\ & & \displaystyle\sum_{t=1}^{\ell} q_t(x_1, \dots, x_{n-1}) g_t(x_1, \dots, x_{n-1}) = 1 \\ & \Longleftrightarrow & \displaystyle\sum_{t=1}^{\ell} q_t(x_1, \dots, x_{n-1}) \sum_{i=1}^{m} \alpha_{it}(x_1, \dots, x_n) f_i(x_1, \dots, x_n) = 1 \\ & \Longleftrightarrow & \displaystyle\sum_{i=1}^{m} p_i(x_1, \dots, x_n) f_i(x_1, \dots, x_n) = 1 \end{array}$$

□

4. Voir à ce sujet les remarques faites après démonstration du théorème 1

Cette démonstration, comme nous avons énoncé au début du paragraphe, nous fournit un algorithme, qui nous permet de réduire le nombre de variables dans un système d'équations polynomiales. De plus, nous pouvons le répéter pour arriver à une seule inconnue

$$\{f_i(x_1,\ldots,x_n)=0\}_{i=1}^m \rightsquigarrow \{g_i(x_1,\ldots,x_{n-1})=0\}_{i=1}^\ell \rightsquigarrow \cdots \rightsquigarrow \{h_i(x_1)=0\}_{i=1}^k$$

Soit $h=\mathrm{pgcd}(h_1,\ldots,h_k)$. Nous pouvons avoir trois cas:

1. $h\equiv 0 \Rightarrow$ infinité de solutions;
2. $h\equiv 1 \Rightarrow$ pas de solution;
3. Si $h\in\mathbb{K}[x_1]$ non constant, alors s'il existe des solutions, elles sont en quantité finie.

Lors de ce dernier point nous pouvons rencontrer un nouveau problème. Il est possible que les solutions numériques entraînent des erreurs d'arrondis, qui peuvent s'agrandir, quand nous remontons le calcul, pour déterminer les inconnues restantes. Pour limiter cela, nous ne remontons pas, mais permutons tout simplement le rôle des x_i dans l'algorithme. Nous arrivons à n systèmes à une inconnue $\{h_{\nu i}(x_i)=0\}_{\nu=1}^{k_i}$, avec $\xi_i^{(j)}$ comme solutions ($i=1,\ldots,n$). Ces $\xi_i^{(j)}$ sont des candidats pour être des solutions du système initial. Nous essayons ensuite, pour chaque i, toutes leurs combinaisons pour voir lesquels sont effectivement les solutions cherchées $(\xi_1^{(j_{t_1})},\ldots,\xi_n^{(j_{t_1})})$.

Exemple 1: Considérons le système suivant:

$$\begin{cases} f_1(x,y)=xy-3x \\ f_2(x,y)=y^2+x \end{cases}$$

et cherchons les valeurs de (x,y) telles que $f_1(x,y)=f_2(x,y)=0$. Si nous éliminons d'abord y, nous trouvons que $x\in\{-9,0\}$. Si nous éliminons d'abord x, nous trouvons que $y\in\{0,3\}$. Nous essayons à ce point tous les couples possibles dans le système, pour voir quels sont ceux qui annulent f_1 et f_2 simultanément. Nous avons

$$\begin{array}{rcccrcr} f_1(-9,0) & = & 27 & \text{et} & f_2(-9,0) & = & -9 \\ f_1(0,0) & = & 0 & \text{et} & f_2(0,0) & = & 0 \\ f_1(-9,3) & = & 0 & \text{et} & f_2(-9,3) & = & 0 \\ f_1(0,3) & = & 0 & \text{et} & f_2(0,3) & = & 9 \end{array}$$

Remarquons que seulement les couples $(0,0)$ et $(-9,3)$ sont solutions. De plus, ils sont les seules solutions, à cause de l'équivalence entre les différents systèmes trouvés par cet algorithme.

Une autre remarque importante est que l'élimination de Kronecker conserve la linéarité, c'est-à-dire que si on part d'un système linéaire, alors tous les systèmes qui suivent sont linéaires aussi. Par contre, dans ce cas, l'élimination de Gauss est beaucoup moins chère. Il faudra donc trouver un autre algorithme plus efficace, ou au moins équivalent à l'élimination de Gauss. Ceci sera l'objet de notre recherche dans les paragraphes qui suivent; (en particulier le résultat sera livré avec l'algorithme 1.4.3 à la page 26).

Théorème 4. ***("Nullstellensatz" de Hilbert, 1893)***
Soient f *et* $f_i \in \mathbb{K}[x_1, \ldots, x_n]$, $i = 1, \ldots, m$. *Supposons que* $f(x_1, \ldots, x_n)$ *s'annule sur toute racine de* $\{f_i(x_1, \ldots, x_n) = 0\}_{i=1}^m$.
Alors il existe des polynômes $p_1, \ldots, p_m \in \mathbb{K}[x_1, \ldots, x_n]$ *et un entier* $k \geq 1$ *tels que*

$$\left(\, f(x_1, \ldots, x_n) \, \right)^k = \sum_{i=1}^{m} p_i(x_1, \ldots, x_n) f_i(x_1, \ldots, x_n)$$

Démonstration. Dans $\mathbb{K}[x_1, \ldots, x_{n+1}]$, considérons le système

$$\begin{cases} f_1(x_1, \ldots, x_n) = 0 \\ \vdots \\ f_m(x_1, \ldots, x_n) = 0 \\ 1 - x_{n+1} f(x_1, \ldots, x_n) = 0 \end{cases}$$

Ce système n'a pas de solutions, car $1 \neq 0$. Par le théorème de Kronecker (théorème 3), $\exists \; \gamma_i \in \mathbb{K}[x_1, \ldots, x_{n+1}]$ tels que

$$\gamma_{m+1}(x_1, \ldots, x_{n+1}) \left(\, 1 - x_{n+1} f(x_1, \ldots, x_n) \, \right) + \sum_{i=1}^{m} \gamma_i(x_1, \ldots, x_{n+1}) f_i(x_1, \ldots, x_n) = 1$$

Cette égalité reste vraie $\forall \; x_{n+1} \in \mathbb{K}$. En particulier, elle est vraie pour

$$x_{n+1} = \frac{1}{f(x_1, \ldots, x_n)}$$

Dans ce cas, nous avons

$$\sum_{i=1}^{m} \gamma_i(x_1, \ldots, x_n, \frac{1}{f(x_1, \ldots, x_n)}) f_i(x_1, \ldots, x_n) = 1 \tag{1.6}$$

Soit k le degré maximal de x_{n+1} dans les γ_i. Nous avons que $f^k \cdot \gamma_i(x_1, \ldots, x_n, 1/f) \in \mathbb{K}[x_1, \ldots, x_n]$ pour chaque $i = 1, \ldots, m$. Soient $p_i = f^k \cdot \gamma_i$, $i = 1, \ldots, m$. L'identité (1.6) devient

$$\sum_{i=1}^{m} p_i(x_1, \ldots, x_n) f_i(x_1, \ldots, x_n) = \left(\, f(x_1, \ldots, x_n) \, \right)^k$$

$\square$

Maintenant nous sommes en mesure de donner une deuxième démonstration de la forme plus générale corollaire 1. Démonstration qui va impliquer l'équivalence entre le théorème 3 et le théorème 4.

Corollaire 2. *Soit $(\mathcal{F})$ le système (1.1). Alors il est suffisant et nécessaire qu'il existe des polynômes $p_1, \ldots, p_m \in \mathbb{K}[x_1, \ldots, x_n]$ tels que*

$$\sum_{i=1}^{m} p_i(x_1, \ldots, x_n) f_i(x_1, \ldots, x_n) = 1$$

pour que $(\mathcal{F})$ n'admette pas de solutions.

Démonstration. La suffisance est évidente et se montre par la contraposée. Pour la nécessité, c'est le *Nullstellensatz de Hilbert* qui nous vient à l'aide.
Supposons que $(\mathcal{F})$ n'admette pas de solutions. Alors le polynôme constant 1 s'annule sur chaque racine de $(\mathcal{F})$. Donc

$$\sum_{i=1}^{m} p_i(x_1, \ldots, x_n) f_i(x_1, \ldots, x_n) = 1^k = 1$$

pour certains $p_1, \ldots, p_m \in \mathbb{K}[x_1, \ldots, x_n]$ et un certain entier $k \geq 1$. $\square$

1.3 Théorème de la base de Hilbert

Par la suite, nous allons utiliser une notation multi-indice. Nous allons donc noter par $\mathbb{K}[x]$ l'anneau des polynômes $\mathbb{K}[x_1, \ldots, x_n]$ et par x^α, où $\alpha = (\alpha_1, \ldots, \alpha_n)$ et $x = x_1 \cdots x_n$, le monôme $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$.

Définition 3. *Soit R un anneau commutatif*[5]*. Un* ***idéal*** *$\mathfrak{a} \subset R$ est un groupe additif qui reste stable par multiplication des éléments de R. Si l'idéal $\mathfrak{a}$ est* ***engendré*** *par un nombre fini d'éléments $a_1, \ldots, a_m \in R$, on dit qu'il est* ***de type fini*** *ou* ***de génération finie*** *et se note $\langle a_1, \ldots, a_m \rangle$. L'ensemble $\{a_1, \ldots, a_m\}$ est un* ***générateur***[6] *de l'idéal.*

Exemple 2: Soient $f_1, \ldots, f_m \in \mathbb{K}[x]$. Alors

$$\mathfrak{a} = \left\{ \; f \in \mathbb{K}[x] \mid f(x) = \sum_{i+1}^{m} p_i(x) f_i(x) \; \right\} = \langle f_1, \ldots, f_m \rangle$$

est un idéal de $\mathbb{K}[x]$.

L'exemple 3 ci-dessous nous montre qu'un générateur d'un idéal n'est pas forcément unique, ni en présentation, ni dans le nombre de ses éléments. Par contre, nous verrons par la suite que, comme $\mathbb{K}$ est un corps, il est toujours possible de trouver un générateur fini à un idéal. Or, en ce qui concerne la recherche des solutions d'un système d'équations

5. En général la condition de commutativité n'est pas demandé. Dans ce cas, il faudra distinguer entre idéal à gauche, à droite et bilatère.

6. Par abus de langage, nous allons appeler ici "générateur" l'ensemble des générateurs.

polynomiales, nous pouvons voir le système donné comme générateur d'un idéal. La propriété d'un générateur de n'être pas forcément unique nous donne l'espoir de pouvoir en trouver un autre plus simple.

Exemple 3: Soient $f_1(x) = x^2-1$, $f_2(x) = x^3-1$ et $g_1(x) = x-1$. Alors $\langle f_1, f_2\rangle = \langle g_1\rangle$. En effet, d'une côté $\langle f_1, f_2\rangle \supset \langle g_1\rangle$, car $f_2(x) - xf_1(x) = g_1(x)$. D'autre part, $\langle f_1, f_2\rangle \subset \langle g_1\rangle$, car $f_1(x) = (x+1)g_1(x)$ et $f_2(x) = (x^2+x+1)g_1(x)$.

Lemme 1. *Soient $f_1, \ldots, f_s$ et $g_1, \ldots, g_r \in \mathbb{K}[x]$. Si $\langle f_1, \ldots, f_s\rangle = \langle g_1, \ldots, g_r\rangle$, alors les racines de $\{f_i(x) = 0\}_{i=1}^{s}$ et de $\{g_j(x) = 0\}_{j=1}^{r}$ sont les mêmes.*

Démonstration. Comme $g_j \in \langle f_1, \ldots, f_s\rangle$, $\forall\ j = 1, \ldots, r$, nous pouvons les écrire sous la forme

$$g_j(x) = \sum_{i=1}^{s} p_{ij}(x) f_i(x)$$

Donc chaque racine de $\{f_i(x) = 0\}_{i=1}^{s}$ est racine de chaque $g_j(x)$. De même $f_i \in \langle g_1, \ldots, g_r\rangle$, donc

$$f_i(x) = \sum_{j=1}^{r} q_{ji}(x) g_j(x)$$

Donc chaque racine de $\{g_j(x) = 0\}_{j=1}^{r}$ est racine de $f_i(x)$ et ceci pour tout $i = 1, \ldots, s$. $\square$

Définition 4. *Un anneau commutatif s'appelle* ***nœthérien*** *si et seulement si toute suite croissante d'idéaux*

$$\mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \cdots \subset \mathfrak{a}_n \subset \cdots$$

admet un plus grand élément, i.e. $\exists\ N \in \mathbb{N}$ t.q. $\mathfrak{a}_N = \mathfrak{a}_{N+\ell}\ \forall\ \ell \in \mathbb{N}$.

Lemme 2. *Soit R un anneau commutatif. Alors R est nœthérien si et seulement si tout idéal $\mathfrak{a} \subset R$ est de génération finie.*

Démonstration. Voyons que si tout idéal de R est de génération finie, alors R est nœthérien. Soit

$$\mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \cdots \subset \mathfrak{a}_n \subset \cdots$$

une suite croissante d'idéaux dans R. Soit

$$\mathfrak{a} = \bigcup_{k \geq 1} \mathfrak{a}_k$$

$\mathfrak{a}$ est un idéal de R. En effet, $\mathfrak{a}$ est non vide, car les $\mathfrak{a}_k$ le sont. De plus, si a et $b \in \mathfrak{a}$, alors $\exists\ k \in \mathbb{N}$ tel que a et $b \in \mathfrak{a}_k$. Donc $-a$, $a+b$ et $\lambda a \in \mathfrak{a}_k \subset \mathfrak{a}$, où $\lambda \in R$. Mais, par hypothèse, tout idéal de R est de génération finie. Donc

$$\mathfrak{a} = \langle a_1, \ldots, a_n \rangle$$

pour certains $a_1, \ldots, a_n$. Or, $\forall\ a_i\ \exists\ k_i$ t.q. $a_i \in \mathfrak{a}_{k_i}$. Soit $N = \max\{k_1, \ldots, k_n\}$. Nous avons que $\mathfrak{a}_N = \mathfrak{a}_{N+\ell}$, $\forall \ell \geq 0$, d'où R est nœthérien.

Réciproquement, nous montrons par l'absurde que si R est nœthérien alors chacun de ses idéaux est de type fini.
En effet, s'il existe un idéal $\mathfrak{a}$ dans R tel que $\mathfrak{a} \neq \langle a_1, \ldots, a_m \rangle$, $\forall m \in \mathbb{N}$ et $\forall a_i \in R$, nous pouvons construire par récurrence une suite croissante d'idéaux qui ne se stabilisera jamais de la façon suivante.
Soit $a_1 \in \mathfrak{a}$. Soit $\mathfrak{a}_1 := \langle a_1 \rangle$. Puis, par récurrence, soient

$$a_k \in \mathfrak{a} \setminus \bigcup_{i=1}^{k-1} \mathfrak{a}_i$$

et

$$\mathfrak{a}_k := \langle a_1, \ldots, a_k \rangle$$

Nous obtenons ainsi la suite

$$\mathfrak{a}_1 \subsetneq \mathfrak{a}_2 \subsetneq \cdots \subsetneq \mathfrak{a}_n \subsetneq \cdots$$

□

Lemme 3. *Si R est un anneau nœthérien, alors $R[x]$ est nœthérien.*

Démonstration. Soit

$$\mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \cdots \subset \mathfrak{a}_n \subset \cdots$$

une suite croissante d'idéaux dans $R[x]$. Soit $\mathfrak{a}_n$ un élément de cette suite. Soit

$$I_{nk} = \{r \in R \mid\ \exists\ f(x) \in \mathfrak{a}_n, \deg f = k\}$$

Affirmation: I_{nk} est un idéal dans R.
En effet, soient $r,\ s \in I_{nk}$. Alors $\exists\ f(x),\ g(x) \in \mathfrak{a}_n$ avec

$$f(x) = rx^k + \cdots$$
$$g(x) = sx^k + \cdots$$

Soit $\lambda \in R$. Alors

$$\begin{array}{ccc} f(x)+g(x) & = & (r+s)x^k + \cdots \quad \in \mathfrak{a}_n \\ \lambda f(x) & = & (\lambda r)x^k + \cdots \quad \in \mathfrak{a}_n \end{array}$$

D'où $r+s$ et $\lambda r \in I_{nk}$.

Nous avons que $I_{nk} \subset I_{n,k+1}$, car $r \in I_{nk} \Leftrightarrow \exists f(x) = rx^k + \cdots \in \mathfrak{a}_n$, donc $xf(x) \in \mathfrak{a}_n$ et $xf(x) = rx^{k+1} + \cdots \Rightarrow r \in I_{n,k+1}$.
De même $I_{nk} \subset I_{n+1,k}$, car $r \in I_{nk} \Leftrightarrow \exists f(x) = rx^k + \cdots \in \mathfrak{a}_n \subset \mathfrak{a}_{n+1} \Rightarrow r \in I_{n+1,k}$.
Ces deux résultats livrent le tableau suivant

$$\begin{array}{ccccccccc}
I_{10} & \subset & I_{20} & \subset & I_{30} & \subset & I_{40} & \cdots \\
\cap & & \cap & & \cap & & \cap & \\
I_{11} & \subset & I_{21} & \subset & I_{31} & \subset & I_{41} & \cdots \\
\cap & & \cap & & \cap & & \cap & \\
I_{12} & \subset & I_{22} & \subset & I_{32} & \subset & I_{42} & \cdots \\
\cap & & \cap & & \cap & & \cap & \\
I_{13} & \subset & I_{23} & \subset & I_{33} & \subset & I_{43} & \cdots \\
\vdots & & \vdots & & \vdots & & \vdots & \cdots
\end{array}$$

L'idée maintenant est de considérer une suite diagonale. Par exemple

$$I_{11} \subset I_{22} \subset I_{33} \subset \cdots$$

Comme R est nœthérien, il existe un entier N tel que $I_{NN} = I_{N+\ell,N+\ell}$, $\forall \ell \in \mathbb{N}$. Donc $I_{ts} = I_{NN}$, $\forall\, t, s \geq N$.

De même, pour les N premières lignes

$$I_{1,s} \subset I_{2,s} \subset I_{3,s} \subset \cdots$$

pour $0 \geq s \geq N-1$, il y a un $M'_s \in \mathbb{N}$ tel que $I_{M'_s,s} = I_{M'_s+\ell,s}$ $(\forall \ell \in \mathbb{N})$.
Et pour les premières N colonnes aussi

$$I_{t,1} \subset I_{t,2} \subset I_{t,3} \subset \cdots$$

pour $1 \geq t \geq N-1$, il y a un $M''_t \in \mathbb{N}$ tel que $I_{t,M''_t} = I_{t,M''_t+\ell}$ $(\forall \ell \in \mathbb{N})$.

Soit maintenant

$$k := \max\{N, M'_0, \ldots, M'_{N-1}, M''_1, \ldots, M''_{N-1}\}$$

Donc $I_{ts} = I_{ks}$ si $t \geq k$, $I_{ts} = I_{tk}$ si $s \geq k$ et, en tout cas, $I_{ts} \subset I_{kk}$ si t ou $s \geq k$.

Affirmation: $\mathfrak{a}_{k+\ell} = \mathfrak{a}_k$ $(\forall\, \ell \in \mathbb{N})$.

Supposons, par l'absurde, que $\mathfrak{a}_{k+\ell} \supsetneqq \mathfrak{a}_k$ pour une certaine valeur de ℓ. Alors $\exists f(x) \in \mathfrak{a}_{k+\ell} \setminus \mathfrak{a}_k$ de degré minimal. Disons $\deg f = s$

$$\begin{aligned}
f(x) = rx^s + \cdots \quad &\Longrightarrow \quad r \in I_{k+\ell,s} \subset I_{ks} \\
&\Longrightarrow \quad \exists\; g(x) = rx^s + \cdots \in \mathfrak{a}_k \subset \mathfrak{a}_{k+\ell} \\
&\Longrightarrow \quad f(x) - g(x) \in \mathfrak{a}_{k+\ell}
\end{aligned}$$

Comme $f(x) \notin \mathfrak{a}_k$, alors $f(x) - g(x) \in \mathfrak{a}_{k+\ell} \setminus \mathfrak{a}_k$. Mais $\deg(f-g) \leq s-1$. Ce qui contredit la minimalité du degré de $f(x)$. □

Théorème 5. ***(théorème de la base de Hilbert)***
Soit $\mathfrak{a}$ un idéal de $\mathbb{K}[x_1, \ldots, x_n]$. Alors $\mathfrak{a}$ est de type fini.

Démonstration. $\mathbb{K}$ est un corps. Donc tout idéal est trivial, i.e. $\langle 0\rangle$ et $\langle 1\rangle$ sont les seuls idéaux. En effet si $\mathfrak{a} \neq \{0\}$ est un idéal de $\mathbb{K}$, alors $\exists\, a \in \mathfrak{a}$, $a \neq 0$, et a est inversible dans $\mathbb{K}$ d'inverse a^{-1}. Mais $a^{-1}a = 1 \in \mathfrak{a}$.
Donc chaque idéal de $\mathbb{K}$ est de génération finie et, par le lemme 2, nous avons que $\mathbb{K}$ est nœthérien. Donc, par le lemme 3, $\mathbb{K}[x_1]$ est aussi nœthérien. De même, par récurrence avec le lemme 3 et via l'isomorphisme $\mathbb{K}[x_1, x_2] \backsimeq \mathbb{K}[x_1][x_2]$, $\mathbb{K}[x_1, x_2]$ est nœthérien, etc. $\mathbb{K}[x_1, \ldots, x_n]$ est donc nœthérien et, par le lemme 2, chacun de ses idéaux est de type fini. $\square$

1.4 Terme ordre et un procédé de réduction

Considérons le système (1.1) et examinons la possibilité de trouver un autre système

$$(\mathcal{G})\begin{cases} g_1(x_1, x_2, \ldots, x_n) = 0 \\ \quad\vdots \\ g_\ell(x_1, x_2, \ldots, x_n) = 0 \end{cases} \tag{1.7}$$

plus simple tel que $\langle f_1, \ldots, f_m\rangle = \langle g_1, \ldots, g_\ell\rangle$.

La méthode proposée ici utilise une généralisation de l'algorithme d'Euclide pour calculer le pgcd$(f_1, \ldots, f_m)$. Avec la notation multi-indice $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, nous avons des $\alpha \in \mathbb{N}^n$. Pour atteindre notre but, nous munissons $\mathbb{N}^n$ d'un ordre.
Fait: pour procéder à une division euclidienne, il ne suffit pas de considérer un ordre total sur $\mathbb{N}^n$. L'exemple 4 nous le montre.

Exemple 4: Considérons sur $\mathbb{N}$ l'ordre

$$0 \prec 2 \prec 1 \prec 3 \prec 4 \prec 5 \prec \cdots \prec n \prec n+1 \prec \cdots$$

Soient

$$f(x) = x^3 + 2x + 3x^2 + 5$$
$$g(x) = x - x^2 - 1$$

Nous avons $\deg f = 3$ et $\deg g = 1$. Pour éliminer le terme dominant de $f(x)$, essayons de faire la division $f(x) = q(x)g(x) + r(x)$, avec $\deg r < \deg g$.

$$\begin{array}{rcll} f(x) - x^2 g(x) & = & x^4 + 2x + 4x^2 + 5 & =: f_1(x) \\ f_1(x) - x^3 g(x) & = & x^5 + x^3 + 2x + 4x^2 + 5 & =: f_2(x) \\ \vdots\ \vdots & & \vdots\ \vdots\ \vdots & \quad\vdots \end{array}$$

Remarquons que l'algorithme ne s'arrête pas avec cet ordre, car le degré des $f_i(x)$ augmente à chaque fois.

Définition 5. *Un ordre "<" sur $\mathbb{N}^n$ s'appelle* ***terme ordre*** *si et seulement si*

1. *l'ordre est total;*
2. $0 < \alpha$ $(\forall \alpha \neq 0)$*;*
3. $\alpha < \beta$ *et* $\gamma \in \mathbb{N}^n \Longrightarrow \alpha + \gamma < \beta + \gamma$ $(\forall \alpha, \beta \in \mathbb{N}^n)$.

Dans l'exemple 4 ci-dessus la troisième propriété n'est pas respectée. Des exemples de terme ordres sont:

– **l'ordre lexicographique**

$$(\alpha_1, \ldots, \alpha_n) < (\beta_1, \ldots, \beta_n) \quad \overset{déf.}{\Longleftrightarrow} \quad \begin{array}{l} \exists k,\ 1 \leq k \leq n,\ \text{t.q } \alpha_j = \beta_j \ (\forall\ j < k) \\ \text{et } \alpha_k < \beta_k \end{array}$$

C'est un ordre total et nous avons

$$1 < x_n < x_n^2 < x_n^3 < \cdots < x_{n-1} < x_{n-1}x_n < x_{n-1}x_n^2 < \cdots < x_{n-1}^2 < \cdots$$

– **l'ordre lexicographique inverse**

$$(\alpha_1, \ldots, \alpha_n) < (\beta_1, \ldots, \beta_n) \quad \overset{déf.}{\Longleftrightarrow} \quad \begin{array}{l} \exists k,\ 1 \leq k \leq n,\ \text{t.q } \alpha_j = \beta_j \ (\forall\ j > k) \\ \text{et } \alpha_k < \beta_k \end{array}$$

C'est un ordre total et nous avons

$$1 < x_1 < x_1^2 < x_1^3 < \cdots < x_2 < x_1x_2 < x_1^2x_2 < \cdots < x_2^2 < \cdots$$

– **l'ordre par degré**

$$(\alpha_1, \ldots, \alpha_n) < (\beta_1, \ldots, \beta_n) \quad \overset{déf.}{\Longleftrightarrow} \quad \begin{array}{l} \displaystyle\sum_{j=1}^n \alpha_j < \sum_{j=1}^n \beta_j \\ \text{ou} \\ \displaystyle\sum_{j=1}^n \alpha_j = \sum_{j=1}^n \beta_j \text{ et } (\alpha_1, \ldots, \alpha_n) < (\beta_1, \ldots, \beta_n) \\ \text{avec l'ordre lexicographique inverse.} \end{array}$$

C'est un ordre total et nous avons

$$1 < x_1 < x_2 < x_3 < \cdots < x_1^2 < x_1x_2 < x_1x_3 < \cdots < x_1^3 < x_1^2x_2 < \cdots$$

Il est possible de trouver d'autres terme ordres lexicographiques, tout simplement en échangeant les rôles des x_i.

Définition 6. *Soit* $f(x_1,\ldots,x_n) = \sum_{finie} c_\alpha x^\alpha$. *Notons par*

$$\deg f = \max\{\alpha \mid \ c_\alpha \neq 0\} \ \ et \ \ tp(f) = c_{\deg f} x^{\deg f}$$

le ***degré*** *de* f *et le* ***terme principal*** *de* f.

Exemple 5: Soit $f(x) = 3x_1^2x_2^5x_3^3 + 4x_1^2x_3^6 + 2 \in \mathbb{R}[x_1,x_2,x_3]$.

- Ordre par degré: $\deg f = (2,5,3)$ et $tp(f) = 3x_1^2x_2^5x_3^3$;
- Ordre lexicographique inverse: $\deg f = (2,0,6)$ et $tp(f) = 4x_1^2x_3^6$.

Théorème 6. ***(Principe du bon ordre)***
Soit "$<$" un terme ordre sur $\mathbb{N}^n$.
Alors

$$\forall\ A \subset \mathbb{N}^n,\ A \neq \emptyset,\ \exists\ \alpha \in A\ \forall\ \beta \in A \setminus \{\alpha\}\ on\ a\ \alpha < \beta$$

Autrement dit, "$<$" est un bon ordre.

Démonstration. (Par l'absurde)
Supposons que $\exists A \neq \emptyset$ tel que $\forall \alpha \in A\ \exists \beta \in A \setminus \{\alpha\} :\ \alpha > \beta$.
Dans ce cas, nous pouvons trouver une suite infinie décroissante dans A

$$\alpha_1 > \alpha_2 > \alpha_3 > \cdots \tag{1.8}$$

Donc $x^{\alpha_{i+1}} \notin \langle x^{\alpha_1},\ldots,x^{\alpha_i}\rangle$. En effet, si $x^{\alpha_{i+1}} \in \langle x^{\alpha_1},\ldots,x^{\alpha_i}\rangle$, alors $\exists\ p_j \in \mathbb{K}[x]$ tels que

$$\begin{aligned} x^{\alpha_{i+1}} &= \sum_{j=1}^{i} p_j(x) x^{\alpha_j} \\ &= \sum_{j,\gamma} c_{j\gamma} x^{\alpha_j+\gamma} \end{aligned}$$

Donc $\exists \gamma \in \mathbb{N}^n$ tel que $\alpha_{i+1} = \alpha_j + \gamma$, pour un certain $j \in \{1,\ldots,i\}$. Mais comme nous sommes confrontés avec un terme ordre, nous avons que $\alpha_{i+1} = \alpha_j + \gamma \geq \alpha_j$. Ce qui est en contradiction avec (1.8).

Nous trouvons ainsi une suite infinie strictement croissante d'idéaux

$$\langle x^{\alpha_1}\rangle \subsetneq \langle x^{\alpha_1}, x^{\alpha_2}\rangle \subsetneq \cdots \subsetneq \langle x^{\alpha_1},\ldots,x^{\alpha_i}\rangle \subsetneq \cdots$$

dans $\mathbb{K}[x]$, contredisant le théorème de la base de Hilbert (théorème 5). $\square$

1.4.1 Procédé de réduction

Nous allons maintenant généraliser l'algorithme d'Euclide avec des polynômes à plusieurs variables.
Soit donc "<" un terme ordre sur $\mathbb{N}^n$ et soit aussi $f \in \mathbb{K}[x_1, \ldots, x_n]$. Nous cherchons des polynômes $g_1, \ldots, g_m$ et $r \in \mathbb{K}[x_1, \ldots, x_n]$ tels que

$$f(x) = \sum_{i=1}^{m} u_i(x) g_i(x) \; + r(x)$$

avec u_i et $r \in \mathbb{K}[x_1, \ldots, x_n]$, $\deg f > \deg r$ et $tp(g_i) \nmid tp(r)$. Ceci est possible s'il existe un i, $1 \leq i \leq m$, tel que $x^{\deg g_i}$ divise $x^{\deg f}$, i.e. $\exists \gamma \in \mathbb{N}^n$ tel que $\deg f = \deg g_i + \gamma$.
Dans ce cas, posons $u_j = 0$, $\forall\, j \neq i$, et $u_i = tp(f)/tp(g_i)$, donc $\deg u_i = \gamma$.
Ce passage est justifié car les systèmes

$$\left\{ \begin{array}{c} f(x_1, \ldots, x_n) = 0 \\ g_1(x_1, \ldots, x_n) = 0 \\ \vdots \\ g_m(x_1, \ldots, x_n) = 0 \end{array} \right. \quad \text{et} \quad \left\{ \begin{array}{c} r(x_1, \ldots, x_n) = 0 \\ g_1(x_1, \ldots, x_n) = 0 \\ \vdots \\ g_m(x_1, \ldots, x_n) = 0 \end{array} \right.$$

sont équivalents [7]. L'algorithme suivant, décrit en pseudo-code, est alors fini. Ceci est dû au fait que nous avons une suite de degrés strictement décroissante.

1.4.2 Algorithme de réduction

Input: $f \in \mathbb{K}[x_1, \ldots, x_n]$ et $g_i \in \mathbb{K}[x_1, \ldots, x_n]$ $(i = 1, \ldots, m)$.

Output: $r \in \mathbb{K}[x_1, \ldots, x_n]$ et $f = \sum_{i=1}^{m} u_i g_i \; + r$ avec $x^{\deg g_i}$ ne divise pas $x^{\deg r}$ $(\forall i = 1, \ldots, m)$.

Code:

$h := f,\ r := 0$
while $(h \not\equiv 0)$ **do**
 if $(\exists\ i : x^{\deg g_i} | x^{\deg h})$ **then**
 $h := h - \frac{tp(h)}{tp(g_i)} g_i$
 else
 $r := r + tp(h),\ h := h - tp(h)$
 end if
end while do

7. Je me permets d'insister sur l'équivalence des deux systèmes, étant donné que c'est la seule chose qui nous assure que nous ne perdons pas de solutions, en passant d'un système à l'autre.

Exemple 6: Voici deux applications différentes de l'algorithme.
Considérons le problème

$$\begin{aligned} f(x,y) &= x^6y^2+3x^2y+1 \\ g_1(x,y) &= xy-3 \\ g_2(x,y) &= y^2-2 \end{aligned}$$

Choisissons et fixons comme terme ordre l'ordre par degré.

Opération	$h(x,y)$	$r(x,y)$	Opération	$h(x,y)$	$r(x,y)$
	$x^6y^2+3x^2y+1$	0		$x^6y^2+3x^2y+1$	0
$-x^5y\cdot g_1$	$3x^5y+3x^2y+1$	0	$-x^6\cdot g_2$	$2x^6+3x^2y+1$	0
$-3x^4\cdot g_1$	$9x^4+3x^2y+1$	0		$3x^2y+1$	$2x^6$
	$3x^2y+1$	$9x^4$	$-3x\cdot g_1$	$9x+1$	$2x^6$
$-3x\cdot g_1$	$9x+1$	$9x^4$		1	$2x^6+9x$
	1	$9x^4+9x$		0	$2x^6+9x+1$
	0	$9x^4+9x+1$			

Nous trouvons deux solutions différentes. Ceci dépend des choix différents que nous avons faits pour arriver au résultat cherché. Cependant les polynômes trouvés sont équivalents et plus simples que la fonction $f(x,y)$ initiale. Les deux solutions sont

$$\begin{aligned} f(x,y)-(x^5y+3x^4+x)\cdot g_1(x,y) &= 9x^4+9x+1 \\ f(x,y)-3x\cdot g_1(x,y)-x^6\cdot g_2(x,y) &= 2x^6+9x+1 \end{aligned}$$

Notation: Soit $G=\{g_1,\ldots,g_m\}$ un ensemble de polynômes. Notons "f est réduit à r modulo G" par $f \stackrel{G}{\longrightarrow} r$.

Définition 7. *Un ensemble $G=\{g_1,\ldots,g_m\}$ est* ***réduit*** *si et seulement si $\forall g_i\in G$ on a $g_i \stackrel{G\setminus\{g_i\}}{\longrightarrow} r \Longrightarrow r=g_i$.*

1.4.3 Algorithme de réduction d'une base

Input: $\mathfrak{a}=\langle f_1,\ldots,f_m\rangle$
avec $m\geq 1$ et $f_j\not\equiv 0$.

Output: une base réduite $\langle g_1,\ldots,g_\ell\rangle$ de $\mathfrak{a}$.

Code:

$g_j := f_j$ ($\forall\ j = 1, \ldots, m$)
$G := \{g_1, \ldots, g_m\}$
$k := 1$
while ($k \leq m$) **do**
  $g_k \xrightarrow{G \setminus \{g_k\}} r$
  if ($r \not\equiv g_k$) **then**
  $g_k := r$, $k := 1$
 else
  $k := k + 1$
  end if
end while do
Enlever les fonctions zéro dans G

Remarque: Cet algorithme est une généralisation de l'algorithme d'Euclide. En effet, si les polynômes ont une variable, c'est l'algorithme d'Euclide qui s'applique. De plus, pour un système linéaire, l'algorithme est équivalent à l'élimination de Gauss sans pivot.

Exemple 7: Soit $\mathfrak{a} = \langle f, g \rangle$ avec $f(x)$ et $g(x)$ polynômes à une variable.
Si, par exemple, $\deg f \geq \deg g$, alors nous divisons $f(x)$ par $g(x)$. Donc $f(x) = q(x)g(x) + r(x)$ avec $\deg r < \deg g$. Puis, nous remplaçons $f(x)$ par $r(x)$ et continuons ainsi, jusqu'à ce que nous arrivions à $\mathfrak{a} = \langle \mathrm{pgcd}(f, g) \rangle$.

Exemple 8: Appliquons l'algorithme au système linéaire suivant à coefficients tous non nuls et distincts.

$$(\mathcal{F}) \left\{ \begin{array}{ll} f_1 : & a_{11}x_1 + a_{12}x_2 + a_{13}x_3 - b_1 = 0 \\ f_2 : & a_{21}x_1 + a_{22}x_2 + a_{23}x_3 - b_2 = 0 \\ f_3 : & a_{31}x_1 + a_{32}x_2 + a_{33}x_3 - b_3 = 0 \end{array} \right.$$

Dans ce cas $m = 3$.
Posons $k := 1$ et $g_i := f_i$ ($i = 1, 2, 3$).

1. $g_1 \xrightarrow{\mathcal{F} \setminus \{g_1\}} r$: Faisons, par exemple, $r := g_1 - \frac{a_{13}}{a_{23}} g_2 = a'_{11}x_1 + a'_{12}x_2 - b'_1$. [8] Nous avons que $\deg r < \deg g_2$ et $\deg r < \deg g_3$, Nous avons donc terminé l'opération de réduction $g_1 \xrightarrow{\mathcal{F} \setminus \{g_1\}} r$. Comme $g_1 \not\equiv r$, nous posons $g_1 := r$ et $k := 1$.

2. Reprenons avec $g_1 \xrightarrow{\mathcal{F} \setminus \{g_1\}} r$. Comme $\deg g_1 < \deg g_2$ et $\deg g_1 < \deg g_3$, nous ne pouvons plus faire d'autres divisions. Donc $g_1 \equiv r$. Posons $k := k + 1 = 2$.

3. $g_2 \xrightarrow{\mathcal{F} \setminus \{g_2\}} r$: Faisons-le en deux étapes.

   (a) $g_2 \xrightarrow{\{g_3\}} r_1$: $r_1 := g_2 - \frac{a_{23}}{a_{33}} g_3 = a'_{21}x_1 + a'_{22}x_2 - b'_2$. On laisse invarié k.

8. J'aurais pu faire aussi $g_1 \xrightarrow{\{g_2\}} r$.

(b) $r_1 \xrightarrow{\{g_1\}} r$: $r := r_1 - \frac{a'_{22}}{a'_{12}} g_1 = a''_{21}x_1 - b''_2$. On laisse invarié k.

Nous avons maintenant terminé le procédé de réduction $g_2 \xrightarrow{\mathcal{F}\setminus\{g_2\}} r$ et, comme $g_2 \not\equiv r$, posons $g_2 := r$ et $k := 1$.
A ce point, le système est triangulaire

$$\begin{cases} g_2: \; a''_{21}x_1 = b''_2 \\ g_1: \; a'_{11}x_1 + a'_{12}x_2 = b'_1 \\ g_3: \; a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3 \end{cases}$$

et nous pourrions interrompre l'exemple ici. Nous voyons clairement la relation avec l'élimination de Gauss.

4. Continuons en reprenant $g_1 \xrightarrow{\mathcal{F}\setminus\{g_1\}} r$ et procédons avec $g_1 \xrightarrow{\{g_2\}} r$.
$r := g_1 - \frac{a'_{11}}{a''_{21}} g_2 = a''_{12}x_2 - b''_1$. Comme $g_1 \not\equiv r$ nous posons $g_1 := r$ et $k := 1$.

5. Comme précédemment $g_1 \xrightarrow{\mathcal{F}\setminus\{g_1\}} r$ avec $r \equiv g_1$. Soit donc $k := k+1 = 2$.

6. De même $g_2 \xrightarrow{\mathcal{F}\setminus\{g_2\}} r$ livre $r \equiv g_2$. Nous posons finalement $k := k+1 = 3$.

7. $g_3 \xrightarrow{\mathcal{F}\setminus\{g_3\}} r$:

   (a) $g_3 \xrightarrow{\{g_1\}} r_1$ donne $r_1 := g_3 - \frac{a_{32}}{a''_{12}} g_1 = a'_{31}x_1 + a'_{33}x_3 - b'_3$.

   (b) $r_1 \xrightarrow{\{g_2\}} r$ donne $r := r_1 - \frac{a'_{31}}{a''_{21}} g_2 = a''_{33}x_3 - b''_3$.

   La réduction $g_3 \xrightarrow{\mathcal{F}\setminus\{g_3\}} r$ est terminée et, comme $g_3 \not\equiv r$, nous posons $g_3 := r$ et $k := 1$.

8. A ce point, l'algorithme reprend dès le début et teste que $g_i \xrightarrow{\mathcal{F}\setminus\{g_i\}} g_i$ pour $i = 1, 2, 3$, puis livre le système

$$\begin{cases} g_1: \; a''_{12}x_2 - b''_1 = 0 \\ g_2: \; a''_{21}x_1 - b''_2 = 0 \\ g_3: \; a''_{33}x_3 - b''_3 = 0 \end{cases}$$

Exemple 9: Considérons l'ordre par degré. Soit $\mathfrak{a} = \langle x^2y - 1, xy^3 - 2, x^2y^2 - xy\rangle$. La base est déjà réduite. Par contre le système

$$\begin{cases} f_1 : x^2y - 1 = 0 \\ f_2 : xy^3 - 2 = 0 \\ f_3 : x^2y^2 - xy = 0 \end{cases}$$

n'a pas de solution. En effet $f_3(x,y) = xy(xy-1) = 0 \iff xy = 0$ ou $xy = 1$. Mais si $xy = 0$, alors $f_1(x,y) = 1 \neq 0$. Et $xy = 1$ implique que $f_1(x,y) = x^2 - 1 = 0$ et que

$f_2(x,y) = y^2 - 2 = 0$. Donc $x^2 = 1$ et $y^2 = 2$. D'où $xy = \pm\sqrt{2} \neq 1$. Contradiction. Donc, par le théorème de Kronecker (théorème 3), nous avons

$$1 = \sum_{i=1}^{3} p_i(x,y) f_i(x,y) \Rightarrow 1 \in \mathfrak{a} \Rightarrow \mathfrak{a} = \langle 1 \rangle$$

Comme la méthode utilisée n'est pas très intéressante, à cause d'un manque d'unicité dans la présentation du système final réduit, lançons-nous à la recherche d'une deuxième, qui permettrait de la conquérir.

1.5 Bases de Gröbner

"Wolfgang Gröbner (1899-1980) a été professeur à Innsbruck et a eu parmi ses disciples les professeurs Bruno Buchberger et Gerhard Wanner. Moi, j'ai suivi son dernier cours, mais à l'époque je n'avais rien compris." (Ernst Hairer, *Phrases célèbres*, 17.11.1999)

Définition 8. *Un générateur G d'un idéal $\mathfrak{a}$ s'appelle **base de Gröbner** si et seulement si $\forall\ f \in \mathfrak{a}$ nous avons $f \xrightarrow{G} 0$.*

Exemple 10: Soit $\mathfrak{a} = \langle x^2y - 1, xy^3 - 2, x^2y^2 - xy \rangle$.
Alors $G = \{g_1 := x^2y - 1, g_2 := xy^3 - 2, g_3 := x^2y^2 - xy\}$ n'est pas une base de Gröbner.
En effet $1 \in \mathfrak{a}$, mais $1 \xrightarrow{G} 1 \neq 0$.
Des bases de Gröbner de $\mathfrak{a}$ sont: $\{1\}$ (qui est réduite), $\{1, g_1\}$, $\{1, g_2, g_3\}$, ...

Lemme 4. *Soit $G = \{g_1, \ldots, g_m\}$ une base de Gröbner d'un idéal $\mathfrak{a} \subset \mathbb{K}[x]$. Soit $f \in \mathbb{K}[x]$. Alors*

$$\left.\begin{array}{l} f \xrightarrow{G} r_1 \\ f \xrightarrow{G} r_2 \end{array}\right\} \Longrightarrow r_1 = r_2$$

Autrement dit: le résultat de l'algorithme de réduction est unique.

Démonstration. $f \xrightarrow{G} r_1$, donc

$$f = \sum_{i=1}^{m} u_i g_i + r_1$$

avec $tp(g_i) \nmid tp(r_1)$ ni aucun des autres termes de r_1 ($\forall\ i$).
De même $f \xrightarrow{G} r_2$, donc

$$f = \sum_{i=1}^{m} v_i g_i + r_2$$

avec $tp(g_i) \nmid tp(r_2)$ ni aucun des autres termes de r_2 ($\forall\ i$).
Donc

$$r_1 - r_2 = \sum_{i=1}^{m} (u_i - v_i) g_i \in \mathfrak{a}$$

Comme G est une base de Gröbner, nous avons $r_1 - r_2 \xrightarrow{G} 0$.
Or, r_1 et r_2 sont réduits modulo G. Donc $r_1 - r_2$ l'est aussi et $r_1 - r_2 \xrightarrow{G} r_1 - r_2$. Ce qui implique, par unicité, que $r_1 = r_2$. □

Curiosité: Si $G = \{g_1, \ldots, g_m\}$ est une base de Gröbner, $f \in \mathbb{K}[x]$ et $f \xrightarrow{G} r$, alors nous pouvons écrire

$$f = \sum_{i=1}^{m} u_i g_i + r$$

avec r qui est unique. Par contre les u_i ne le sont pas. Nous pouvons donc trouver des v_i, qui satisfont la même égalité. Si nous faisons la différence $u_i - v_i =: p_i$ nous trouvons une identité $\sum_{i=1}^{m} p_i g_i = 0$ qui est non triviale.
Prenons par exemple $G = \{x^2 + 2y, xy, y^2\}$ et considérons l'ordre par degré. G est une base de Gröbner réduite (admettons-le pour l'instant). Soit

$$f(x, y) = xy^2 + x^2 y + y^2$$

un polynôme de $\mathfrak{a} = \langle G \rangle$. Alors nous pouvons écrire

$$\begin{aligned} f(x,y) &= (x+y) \cdot g_2 + g_3 && (1.9)\\ &= (x+1) \cdot g_3 + x \cdot g_2 && (1.10)\\ &= y \cdot g_1 + y \cdot g_2 - g_3 && (1.11) \end{aligned}$$

Si nous faisons (1.9) − (1.10) nous obtenons l'identité suivante:

$$y \cdot g_2 - x \cdot g_3 = 0$$

Calculons maintenant (1.11) − (1.10). Nous obtenons l'identité suivante:

$$y \cdot g_1 - x \cdot g_2 - 2g_3 = 0$$

etc.
Sans entrer dans les détails, il vaut mieux souligner que certains algorithmes sophistiqués utilisent la propriété de pouvoir trouver des identités non triviales $\sum p_i g_i = 0$, appelées syzygie[9], pour livrer des bases de Gröbner assez rapidement.

Théorème 7. ***(Critère pour une base de Gröbner)***
Soit $G = \{g_1, \ldots, g_m\}$ un générateur d'un idéal $\mathfrak{a} \subset \mathbb{K}[x_1, \ldots, x_n]$. Alors G est une base de Gröbner si et seulement si $\forall\ f \in \mathfrak{a}\ \exists\ i$ tel que $tp(g_i)$ divise $tp(f)$.

9. Nom féminin. Du grec “suzzugia”, union. ASTRON. Conjonction ou opposition de la lune avec le soleil (nouvelle ou pleine lune). – *Le Petit Larousse* , 1993 –

Démonstration. Soit $f \in \mathfrak{a}$. Supposons que G soit une base de Gröbner. Alors $f \xrightarrow{G} 0$. Donc $\exists\ i,\ \alpha \in \mathbb{N}^n$ et $c \in \mathbb{K}$ tels que $tp(f) = cx^\alpha tp(g_i)$, car sinon nous ne pouvons pas réduire f.
Réciproquement, faisons $f \xrightarrow{G} h$. Supposons par l'absurde que $h \not\equiv 0$. Nous avons que h est réduit et aussi que $h \in \mathfrak{a}$, car

$$h = \underbrace{f}_{\in \mathfrak{a}} - \underbrace{\text{comb. lin. des } g_i}_{\in \mathfrak{a}}$$

Donc, par hypothèse, $\exists\ i$ tel que $tp(h) = cx^\alpha tp(g_i)$. D'où h n'est pas réduit et par conséquent $h \equiv 0$. $\square$

Interprétation géométrique. Soient $\mathfrak{a} \subset \mathbb{K}[x_1, \ldots, x_n]$ un idéal et $G = \{g_1, \ldots, g_m\}$ une base de Gröbner de cet idéal.
Fixons un terme ordre sur $\mathbb{N}^n$. Soit $E = \{\deg f \in \mathbb{N}^n ; f \in \mathfrak{a}, f \not\equiv 0\}$. Soit aussi

$$f(x) = \sum_\alpha c_\alpha x^\alpha \in \mathfrak{a}$$

avec $\beta = \deg f$. Si $\gamma = (\gamma_1, \ldots, \gamma_n) \geq 0$ alors

$$x^\gamma f(x) = \sum_\alpha c_\alpha x^{\alpha+\gamma}$$

et $\deg x^\gamma f(x) = \beta + \gamma \geq \beta$.
Ceci se traduit dans la figure 1, pour $n = 2$ et $G = \{x_1^2 x_2^3,\ x_1^4 x_2\}$, avec $E =$ *région hachurée*.

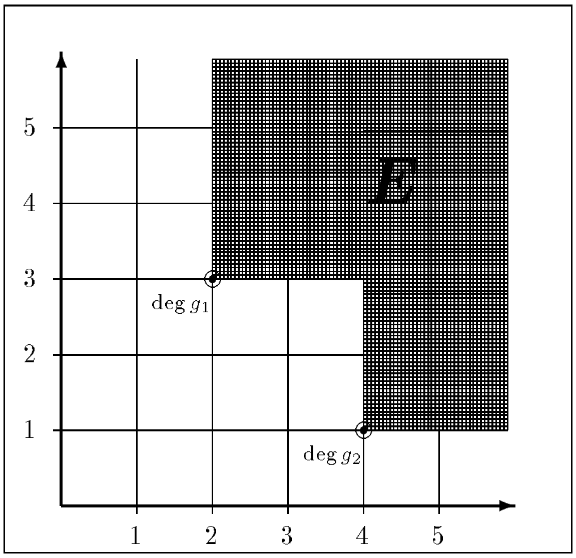

FIGURE 1. *Représentation d'un ensemble E pour $n = 2$.*

Ce résultat se généralise en: $\forall$ coin $\alpha \in E\ \exists\ g_i \in G$ tel que $\deg g_i = \alpha$.

Les deux théorèmes suivants nous donnent l'existence d'une base de Gröbner et l'unicité si elle est réduite. En particulier, la démonstration du théorème 8 nous informe sur la marche à suivre pour la recherche d'une base de Gröbner.

Théorème 8. ***(Existence d'une base de Gröbner)***
Chaque idéal $\mathfrak{a}$ de $\mathbb{K}[x]$ possède une base de Gröbner.

Démonstration. Soit $E = \{\deg f |\ f \in \mathfrak{a},\ f \not\equiv 0\}$. Soient $\alpha_1, \ldots, \alpha_m$ tous les coins distincts de E. Soient aussi $g_1, \ldots, g_m \in \mathfrak{a}$ tels que $tp(g_i) = x^{\alpha_i}$. Alors $G = \{g_1, \ldots, g_m\}$ est une base de Gröbner.
En effet, $f \in \mathfrak{a} \ \Rightarrow\ \exists\ i$ tel que $tp(f) = cx^{\gamma_i} tp(g_i)$ pour un certain $\gamma_i \in \mathbb{N}^n$ et $c \in \mathbb{K}$.
Voyons que G est un générateur de $\mathfrak{a}$.
Soit $f \in \mathfrak{a}$. Faisons $f \xrightarrow{G} r$. Nous avons r qui est réduit. Ecrivons

$$r = f - \sum_{i=1}^{m} u_i g_i \ \in \mathfrak{a}$$

Supposons par l'absurde que $r \not\equiv 0$. Alors $\deg r \in E$. Donc $\exists\ i$ tel que $tp(r) = cx^{\gamma} tp(g_i)$ pour un certain $\gamma_i \in \mathbb{N}^n$ et $c \in \mathbb{K} \setminus \{0\}$, ce qui impliquerait que r n'était pas réduit.
Donc $r \equiv 0$. D'où $f = \sum u_i g_i$ et $\mathfrak{a} \subset \langle G \rangle$. Mais $G \subset \mathfrak{a}$, donc $\mathfrak{a} \supset \langle G \rangle$, d'où $\mathfrak{a} = \langle G \rangle$. ☐

Il peut y avoir plusieurs bases de Gröbner, mais si nous ne considérons que les bases réduites, alors nous avons l'unicité à une constante multiplicative près. Le théorème suivant le prouve.

Théorème 9. ***(Unicité d'une base de Gröbner réduite)***
Fixons un terme ordre sur $\mathbb{K}[x]$. Soient $G = \{g_1, \ldots, g_m\}$ et $F = \{f_1, \ldots, f_s\}$ deux bases de Gröbner réduites du même idéal $\mathfrak{a} \subset \mathbb{K}[x]$. Supposons que le terme principal de chaque g_i et de chaque f_j soit 1. Alors $F = G$.

Démonstration. Soit $g_i \in G \subset \mathfrak{a}$. Comme F est une base de Gröbner, par le théorème 7, nous avons $tp(g_i) = x^{\alpha} tp(f_j)$ pour un certain j et un certain $\alpha \in \mathbb{N}^n$.
De même, $tp(f_j) = x^{\beta} tp(g_k)$. Donc $tp(g_i) = x^{\alpha+\beta} tp(g_k)$.
Comme G est réduite, nous avons $i = k$ et $\alpha + \beta = 0$. Mais $\alpha, \beta \geq 0$, donc $\alpha = \beta = 0$ et $\forall\ i\ \exists\ j$ tel que $tp(g_i) = tp(f_j)$.
Le fait que F est réduite implique qu'un tel j est unique. Donc $s \geq m$.
De même, nous montrons que $\forall\ i\ \exists! j$ tel que $tp(g_i) = tp(f_j)$. Ceci implique que $s \leq m$, d'où l'égalité.
Nous avons donc une bijection $\sigma : G \to F$ telle que $tp(\sigma(g_i)) = tp(g_i)$. Reste à voir que $f_j := \sigma(g_i) = g_i$.
Il est évident que $g_i - f_j \xrightarrow{G} 0$.
Supposons $f_j \neq g_i$ et soit $c_\alpha x^\alpha = tp(g_i - f_j)$. Nous avons $c_\alpha \neq 0$ et $\alpha < \deg g_i = \deg f_j$.
Mais $g_i - f_j \in \mathfrak{a} \ \Rightarrow\ \exists\ k$ et β tels que $x^\alpha = x^\beta tp(g_k)$.
Donc, si x^α se trouve dans g_i, alors G ne serait pas réduite, car nous pourrions réduire g_i grâce à g_k. Et si x^α est dans f_j alors f_j ne serait pas réduit non plus. Contradiction. ☐

1.6 Construction d'une base de Gröbner

Inspirés par l'interprétation géométrique du théorème 7, nous allons construire une base de Gröbner en recherchant les coins de l'ensemble des degrés des polynômes.

Exemple 11: Soient $f_1(x,y) = x^3y^4 - 3x$ et $f_2(x,y) = x^4y^2 + 2y$.
Posons $\mathfrak{a} = \langle f_1, f_2 \rangle$ et cherchons une base de Gröbner pour $\mathfrak{a}$ selon l'ordre par degré.
Nous avons

$$xf_1 - y^2 f_2 = \underbrace{-2y^3 - 3x^2}_{:=f_3} \in \mathfrak{a}$$

Remarquons que maintenant nous pouvons réduire f_1.

$$f_1 \xrightarrow{\{f_3\}} f_4 = f_1 + \frac{1}{2}x^3 y f_3 = -x^5 y - 3x$$

etc.

La figure 2 nous montre que f_1 n'est pas un coin de l'ensemble E rencontré dans le chapitre précédent. De plus, nous avons trouvé à ce stade deux nouveaux coins: f_3 et f_4.

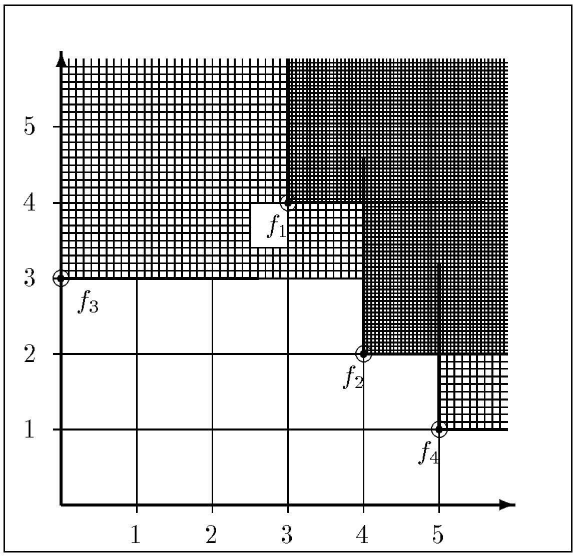

FIGURE 2.

Définition 9. *Soient f et $g \in \mathbb{K}[x]$ non nuls. Soit $x^\gamma = \mathrm{ppcm}\{x^{\deg f},\ x^{\deg g}\}$.*
Le ***polynôme de soustraction*** *est défini par*

$$S(f,g) = x^{\gamma - \deg f} f(x) - c\ x^{\gamma - \deg g} g(x)$$

où c est tel que x^γ disparaît.

Remarque: Soit $\mathfrak{a}$ un idéal. Alors f_i et $f_j \in \mathfrak{a} \;\Rightarrow\; S(f_i, f_j) \in \mathfrak{a}$.

Lemme 5. *Soit $G = \{g_1, \ldots, g_m\} \subset \mathbb{K}[x]$ tel que $\deg g_k = \delta$ pour tout $k = 1, \ldots, m$. Soit*

$$\ell(x) = \sum_{k=1}^{m} c_k g_k(x)$$

avec $\deg \ell < \delta$. Alors nous avons $\sum_{k=1}^{m} c_k = 0$ et nous pouvons écrire

$$\ell = \sum_{1 \leq i < j \leq m} d_{ij} S(g_i, g_j)$$

Démonstration. Supposons $tp(g_k) = x^\delta$. Nous avons le droit de le faire, car, dans le cas contraire, il suffit de diviser les g_k par leurs coefficients dominants et les utiliser pour multiplier les c_k relatifs. Nous avons donc $S(g_i, g_j) = g_i - g_j$ et $\deg S(g_i, g_j) < \delta$. Maintenant, à l'aide d'une sommation partielle d'Abel, nous écrivons

$$\begin{aligned}
\sum_{k=1}^{m} c_k g_k \;\; = \;\; & c_1(g_1 - g_2) + \\
& +(c_1 + c_2)(g_2 - g_3) + \\
& +(c_1 + c_2 + c_3)(g_3 - g_4) + \cdots \\
& \cdots + \underbrace{(c_1 + \cdots + c_m)}_{=0, \text{ car } \deg \ell \,<\, \delta} g_m
\end{aligned}$$

Ce qui nous livre justement

$$\ell = \sum_{1 \leq i < j \leq m} d_{ij} S(g_i, g_j)$$

avec

$$d_{ij} = \begin{cases} 0 & \text{si } i \neq j - 1 \\ c_1 + \cdots + c_i & \text{sinon} \end{cases}$$

□

La démonstration du théorème suivant, dû à Buchberger, est très importante, puisqu'elle est constructive, c'est-à-dire qu'elle nous livre une méthode pour calculer une base de Gröbner.

Théorème 10. *(Buchberger, 1965)*
Soit $G = \{f_1, \ldots, f_m\}$ un générateur d'un idéal $\mathfrak{a}$ dans $\mathbb{K}[x]$. Alors G est une base de Gröbner si et seulement si $\forall\ f, g \in G$ nous avons que $S(f, g) \xrightarrow{G} 0$.

Démonstration. La nécessité est évidente, car si f et $g \in G$, alors $S(f,g) \in \mathfrak{a}$, donc $S(f,g) \xrightarrow{G} 0$ par définition de base de Gröbner.
Pour la suffisance, soit $h \in \mathfrak{a}$. A voir $h \xrightarrow{G} 0$.
Supposons h réduit et, par l'absurde, $h \not\equiv 0$. Ecrivons

$$h(x) = \sum_{i,\alpha} c_{i\alpha} x^\alpha f_i(x)$$

Soit $\delta = \max\{\alpha + \deg f_i \mid \ c_{i\alpha} \neq 0\}$. Nous allons démontrer que $h(x)$ possède une autre représentation:

$$h(x) = \sum_{i,\alpha} \widehat{c_{i\alpha}} x^\alpha f_i(x)$$

où $\hat{\delta}_1 = \max\{\alpha + \deg f_i \ \mid \ \widehat{c_{i\alpha}} \neq 0\}$ est tel que $\hat{\delta} < \delta$ pour l'ordre choisit. Nous serions donc en mesure de construire une suite infinie strictement décroissante

$$\delta > \delta_1 > \delta_2 > \cdots$$

ce qui est impossible. D'où $h \equiv 0$.

Ecrivons dans ce but $h(x) = \ell(x) + r(x)$, où chaque terme de ℓ a un degré de δ et les autres termes de $h(x)$ se trouvent dans $r(x)$. Nous avons donc

$$\ell(x) = \sum_{\alpha + \deg f_i = \delta} c_{i\alpha} x^\alpha f_i(x)$$

Il est évident que $\deg \ell \leq \delta$.
Affirmation: $\deg \ell < \delta$.
En effet, si $\deg \ell = \delta$ alors h ne serait pas réduit.
Par le lemme 5, nous avons

$$\ell(x) = \sum_{\substack{\alpha + \deg f_i = \delta \\ \beta + \deg f_j = \delta}} d_{i\alpha j\beta} S(x^\alpha f_i, x^\beta f_j) \tag{1.12}$$

Etudions cette égalité. Nous pouvons encore supposer que les f_k sont unitaires. Soit $x^{\gamma_{ij}} = \mathrm{ppcm}\{tp(f_i), tp(f_j)\}$. Nous avons ainsi $\alpha + \deg f_i = \beta + f_j = \delta$, donc $x^{\gamma_{ij}}$ divise x^δ, car x^δ est un multiple commun de $tp(f_i)$ et $tp(f_j)$ et $x^{\gamma_{ij}}$ est le plus petit commun multiple. Nous avons

$$\begin{aligned}
S(x^\alpha f_i, x^\beta f_j) &= x^\alpha f_i(x) - x^\beta f_j(x) \\
&= x^{\delta - \deg f_i} f_i(x) - x^{\delta - \deg f_j} f_j(x) \\
&= x^{\delta - \gamma_{ij}} x^{\gamma_{ij} - \deg f_i} f_i(x) - x^{\delta - \gamma_{ij}} x^{\gamma_{ij} - \deg f_j} f_j(x) \\
&= x^{\delta - \gamma_{ij}} (x^{\gamma_{ij} - \deg f_i} f_i(x) - x^{\gamma_{ij} - \deg f_j} f_j(x)) \\
&= x^{\delta - \gamma_{ij}} S(f_i, f_j)
\end{aligned}$$

Donc (1.12) peut s'écrire

$$\ell(x) = \sum_{\substack{\alpha + \deg f_i = \delta \\ \beta + \deg f_j = \delta}} d_{i\alpha j\beta} x^{\delta - \gamma_{ij}} S(f_i, f_j) \tag{1.13}$$

Comme f_i et $f_j \in G$, nous avons par hypothèse que $S(f_i, f_j) \xrightarrow{G} 0$. Autrement dit, il existe des $\eta_{ij} \in \mathbb{N}^n$ et des $e_{ijk\eta_{ij}} \in \mathbb{K}$, qui satisfont

$$S(f_i, f_j) = \sum_{k,\eta_{ij}} e_{ijk\eta_{ij}} x^{\eta_{ij}} f_k(x)$$

avec $\eta_{ij} + \deg f_k \leq \deg S(f_i, f_j)$. De plus $\deg S(f_i, f_j) < \gamma_{ij}$. Donc nous pouvons écrire

$$\ell(x) = \sum_{i,j,k} \widehat{c_{ijk}} x^{\delta - \gamma_{ij} + \eta_{ij}} f_k(x)$$

et $\hat{\delta} = \max\{\delta - \gamma_{ij} + \eta_{ij} + \deg f_k : \widehat{c_{ijk}} \neq 0\}$. Mais

$$\delta - \gamma_{ij} + \underbrace{\eta_{ij} + \deg f_k}_{<\gamma_{ij}} < \delta - \gamma_{ij} + \gamma_{ij} = \delta$$

$\forall\ i$, j et k. Donc aussi pour le max, i.e. $\hat{\delta} < \delta$. □

Exemple 12: Soit $\mathfrak{a} = \langle f_1, f_2 \rangle$, où $f_1(x, y) = xy - x$ et $f_2(x, y) = x^2 - y$. Considérons l'ordre par degré. Alors

$$S(f_1, f_2) = x f_1 - y f_2 = -x^2 + y^2 \xrightarrow{\{f_2\}} f_3(x, y) := y^2 - y$$

et $f_3 \in \mathfrak{a}$. Considérons $G = \{f_1, f_2, f_3\}$ et vérifions si G est une base de Gröbner.
1. $S(f_1, f_2) = f_3 \xrightarrow{G} 0$;
2. $S(f_1, f_3) = y f_1 - x f_3 = 0 \xrightarrow{G} 0$;
3. $S(f_2, f_3) = y^2 f_2 - x^2 f_3 = -y^3 + x^2 y = -y f_2 \xrightarrow{G} 0$.

1.6.1 Algorithme de Buchberger pour la construction d'une base de Gröbner

Input: $F = \{f_1, \ldots, f_m\}$, $f_i \neq 0$, générateur d'un idéal $\mathfrak{a} \subset \mathbb{K}[x]$.

Output: $G = \{g_1, \ldots, g_\mu\}$ une base de Gröbner de $\mathfrak{a}$.

Code:

$G := F$ et $B := \{(f_i, f_j) \mid i < j\}$
while $B \neq \emptyset$ **do**

choisir $(u, v) \in B$ et calculer $S(u, v) \xrightarrow{G} h$
if $h \not\equiv 0$ **then**
$B := B \cup \{(w, h) \mid w \in G\}$
$G := G \cup \{h\}$
end if
$B := B \setminus \{(u, v)\}$
end while do

Théorème 11. *L'algorithme de Buchberger pour le calcul d'une base de Gröbner se termine après une nombre fini d'itérations.*

Malheureusement, le théorème ne nous dit pas combien de temps va prendre une application de cet algorithme.

Démonstration. Supposons, par l'absurde, qu'il faut ajouter une infinité de fois un élément h à la base G. Nous obtiendrons une suite strictement croissante

$$G = G_1 \subsetneq G_2 \subsetneq G_3 \subsetneq G_4 \subsetneq \cdots$$

du même idéal. Considérons $\mathfrak{a}_i = \langle tp(g);\ g \in G_i \rangle$. Nous aurions alors la suite d'idéaux

$$\mathfrak{a}_1 \subsetneq \mathfrak{a}_2 \subsetneq \mathfrak{a}_3 \subsetneq \mathfrak{a}_4 \subsetneq \cdots$$

car $G_{i+1} = G_i \cup \{h\}$ avec $tp(h) \notin \mathfrak{a}_i$.
Ceci est en contradiction avec le théorème 5. □

1.7 Résolution d'un système polynomial

Considérons le système (1.1) noté par $(\mathcal{F})$. Soit $\mathfrak{a} = \langle f_1, \ldots, f_m \rangle$. Supposons que pour cet idéal nous connaissions une base de Gröbner réduite $G = \{g_1, \ldots, g_\mu\}$. Il est évident que connaître toutes les solutions de G, s'il y en a, correspond à connaître toutes les solutions de $(\mathcal{F})$, et qu'elles sont les mêmes.

Théorème 12. *Soit G une base de Gröbner de l'idéal $\langle f_1, \ldots, f_m \rangle$, où les f_i sont les polynômes d'un système $(\mathcal{F})$. Alors le système $(\mathcal{F})$ n'a pas de solutions dans $\mathbb{C}$ $\iff$ $1 \in G$.*

Démonstration. La suffisance est évidente, car $1 \neq 0$ dans $\mathbb{C}$.
La nécessité est une conséquence du théorème 3, car

$$1 = \underbrace{\sum_{i=1}^{m} u_i f_i}_{\in\ \mathfrak{a}} \xrightarrow{G} 0 \Rightarrow 1 \in G$$

□

Lemme 6. *Considérons* $f_1(x), \ldots, f_m(x) \in \mathbb{K}[x]$. *Soit* $\mathfrak{a} = \langle f_1(x), \ldots, f_m(x) \rangle$. *Soit* $E = \{\deg h \mid h \in \mathfrak{a}\}$. *Supposons aussi que pour un terme ordre choisi* $\mathbb{N}^n \setminus E$ *est fini.*
Alors $\forall\ x_i\ \exists\ p_i \in \mathbb{K}[x_i] \setminus \{0\}$ *tel que* $\forall\ \hat{x} = (\hat{x}_1, \ldots, \hat{x}_n)$ *solution de* $(\mathcal{F})$ *nous avons* $p_i(\hat{x}_i) = 0,\ (i = 1, \ldots, n)$.

Démonstration. Soit $\mathbb{N}^n \setminus E = \{\alpha_1, \ldots, \alpha_\ell\}$. Soient G une base de Gröbner de l'idéal $\mathfrak{a}$ et $g \in \mathbb{K}[x_1, \ldots, x_n]$. Inspirés par

$$g \xrightarrow{G} h = \sum_{i=1}^{\ell} c_i x^{\alpha_i}$$

où les c_i sont uniques, car G est une base de Gröbner, écrivons

$$\begin{pmatrix} 1 \\ x_1 \\ x_1^2 \\ \vdots \\ x_1^\ell \end{pmatrix} \xrightarrow{G} \underbrace{\begin{pmatrix} c_{01} & c_{02} & \cdots & c_{0\ell} \\ c_{11} & c_{12} & \cdots & c_{1\ell} \\ \vdots & \vdots & & \vdots \\ c_{\ell 1} & c_{\ell 2} & \cdots & c_{\ell\ell} \end{pmatrix}}_{=:C} \begin{pmatrix} x^{\alpha_1} \\ x^{\alpha_2} \\ x^{\alpha_3} \\ \vdots \\ x^{\alpha_\ell} \end{pmatrix} \tag{1.14}$$

La matrice C a $\ell + 1$ lignes et ℓ colonnes. Ceci implique qu'il y a une ligne linéairement dépendante des autres. Il existe donc un vecteur propre $d = (d_0, \ldots, d_\ell)^T$ tel que $d^T C = 0$. Soit

$$p_1(x_1) = \sum_{i=1}^{\ell} d_i x_1^i$$

Nous avons $p_1 \in \mathbb{K}[x_1]$ et $p_1 \xrightarrow{G} 0$. Donc $p_1 \in \mathfrak{a}$ et, par conséquent, s'annule sur tout $\hat{x}_1$. Nous pouvons faire la même construction pour les autres variables. □

Théorème 13. *Soit* G *une base de Gröbner de* $\mathfrak{a} = \langle f_1, \ldots, f_m \rangle$.
Soit aussi $E = \{\deg f \mid f \in \mathfrak{a}\}$. *Alors le système* $(\mathcal{F})$ *possède un nombre fini de solutions si et seulement si:*

1. $\mathbb{N}^n \setminus E$ *est fini;*
2. E *touche tous les axes;*
3. $\forall\ i \in \{1, \ldots, n\}\ \exists\ g \in G$ *tel que* $tp(g) = c_i x_i^{m_i}$, *avec* c_i *constant et* $m_i \geq 0$ *entier.*

Démonstration.
Nécessité: Soient $x^{(1)}, \ldots, x^{(\ell)}$ toutes les solutions dans $\mathbb{C}$ de $(\mathcal{F})$. Posons

$$p(x) = (x_1 - x_1^{(1)})(x_1 - x_1^{(2)}) \cdots (x_1 - x_1^{(\ell)})$$

Or, G est une base de Gröbner dans $\mathbb{K}[x]$, donc aussi dans $\mathbb{C}[x]$ et $p \in \mathbb{C}[x]$.
Soit $\hat{x}$ une solution de $(\mathcal{F})$. Alors $p(\hat{x}) = 0$.
Par le théorème 4, $\exists\ k \in \mathbb{N}$ et $u_i \in \mathbb{K}[x]$ tels que $p^k = \sum u_i f_i \in \mathfrak{a}$.
Donc $\exists\ g \in G$ tel que $tp(p^k) = c x^\alpha tp(g) \iff x_1^{k\ell} = c x^\alpha tp(g)$. D'où $tp(g) = c x_1^{m_1}$.
On repète pour $x_2, \ldots, x_n$.
Suffisance: Supposons $\mathbb{N}^n \setminus E$ fini. Si $(\mathcal{F})$ n'a pas de solution, nous avons terminé. Sinon, soit $\hat{x}$ une solution de $(\mathcal{F})$. Alors, par le lemme 6, $\forall\ x_i\ \exists\ p_i(x_i) \not\equiv 0$ polynôme tel que $p_i(\hat{x}_i) = 0$. Mais chaque p_i n'a qu'un nombre fini de racines. Donc $(\mathcal{F})$ possède un nombre fini de solutions. □

1.7.1 Algorithme A

1. Fixer un terme ordre et calculer une base de Gröbner.

2. Pour chaque variable x_i caluler le polynôme p_i du lemme 6.

3. Trouver toutes les racines des p_i, pour $i = 1, \ldots, n$, en utilisant si c'est nécessaire des méthodes numériques.

4. Tester les différentes combinaisons de telles racines pour voir lesquelles sont des solutions.

Exemple 13: Considérons le système suivant:

$$(\mathcal{F}) \left\{ \begin{array}{rcl} f_1(x,y) & = & xy - 3x \\ f_2(x,y) & = & y^2 - x \\ f_3(x,y) & = & x^2 - 9x \end{array} \right.$$

$G = \{f_1, f_2, f_3\}$ est déjà une base de Gröbner pour l'ordre par degré. La recherche de $p_1(x)$ est facile. En effet, soit

$$p_1(x) = x^2 - 9x$$

Ses racines sont 0 et 9. Pour $p_2(y)$ nous allons utiliser le lemme 6.

$$\begin{pmatrix} 1 \\ y \\ y^2 \\ y^3 \end{pmatrix} \xrightarrow{G} \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 3 & 0 \end{pmatrix}}_{=:\, C} \begin{pmatrix} 1 \\ x \\ y \end{pmatrix}$$

Nous retrouvons la troisième ligne de la matrice C en faisant $y^3 \xrightarrow{G} xy \xrightarrow{G} 3x$. Cherchons un vecteur d tel que $d^T C = 0$. Nous obtenons, par exemple,

$$d = \begin{pmatrix} 0 \\ 0 \\ -3 \\ 1 \end{pmatrix}$$

qui livre le polynôme

$$p_2(y) = y^3 - 3y^2$$

dont les racines sont 0 et 3. Les candidats pour la solution de $(\mathcal{F})$ sont donc: $(0,0)$, $(0,3)$, $(9,0)$ et $(9,3)$. Nous vérifions enfin, à la main, que seulement $(0,0)$ et $(9,3)$ sont solutions de $(\mathcal{F})$. Elles sont les seules.

1.7.2 Algorithme B

1. Fixer un ordre lexicographique, par exemple l'ordre lexicographique inverse.

2. Calculer une base de Gröbner. Elle donne un système de la forme

$$\begin{cases} g_1(x_1) = 0 \\ g_2(x_1, x_2) = 0 \\ \quad\vdots \\ g_{n_2}(x_1, x_2) = 0 \\ g_{n_2+1}(x_1, x_2, x_3) = 0 \\ \quad\vdots \\ g_{n_3}(x_1, x_2, x_3) = 0 \\ g_{n_3+1}(x_1, x_2, x_3, x_4) = 0 \\ \quad\vdots \end{cases}$$

3. Calculer les racines $\hat{x}_1$ de g_1.

4. $\forall\ \hat{x}_1$ caluler $q_{\hat{x}_1}(x_2) = \text{pgcd}\{g_2(\hat{x}_1, x_2), \ldots, g_{n_2}(\hat{x}_1, x_2)\}$.

5. Calculer les racines $\hat{x}_2$ de $q_{\hat{x}_1}$.

6. Repéter 4 et 5 jusqu'à ce qu'on ait trouvé toutes les solutions du système.

Exemple 14: $\{xy - 3x, y^2 - x, x^2 - 9x\}$ est déjà une base de Gröbner pour l'ordre lexicographique inverse.

1. $\hat{x}_1 = 0$ et $\hat{x}_2 = 9$ sont les solutions de $x^2 - 9x = 0$.

2. (a) Pour $\hat{x}_1 = 0$ nous avons avec les deux autres polynômes $0 = 0$ et $y^2 = 0$. Il en résulte que $(\hat{x}_1, \hat{y}_1) = (0, 0)$ est une solution du système.
   (b) Pour $\hat{x}_2 = 9$ nous avons $\text{pgcd}\{9y - 27, y^2 - 9\} = y - 3$. Ce qui livre $(\hat{x}_2, \hat{y}_2) = (9, 3)$ comme deuxième et dernière solution.

Remarque: L'avantage de l'algorithme B consiste à nous permettre de calculer avec des polynômes ayant un petit degré. Le désavantage est que si $\hat{x}_1$ a été obtenu numériquement, les coefficients des polynômes $g_i(\hat{x}_1, x_2) = 0$ seront erronés [10].

1.7.3 Algorithme C

Cet algorithme repose sur la même idée que l'algorithme A. Seulement, ici nous allons rechercher des valeurs propres plutôt que des polynômes à une variable.

10. cf. § 1.9 pour un étude des systèmes polynomiaux perturbés.

Considérons une base de Gröbner G pour un terme ordre choisi et $f \in \mathbb{K}[x]$. Soit $E = \{\deg h \mid h \in \langle G \rangle\}$. Faisons

$$f \xrightarrow{G} r = \sum_{i=1}^{\ell} c_{\alpha_i} x^{\alpha_i}$$

où $\{\alpha_1, \ldots, \alpha_\ell\} \subset \mathbb{N}^n \setminus E$.
Nous remarquons que nous pouvons faire la même chose avec

$$x_i \begin{pmatrix} x^{\alpha_1} \\ \vdots \\ x^{\alpha_\ell} \end{pmatrix} \xrightarrow{G} \begin{pmatrix} & & \\ & C_i & \\ & & \end{pmatrix} \begin{pmatrix} x^{\alpha_1} \\ \vdots \\ x^{\alpha_\ell} \end{pmatrix}$$

où C_i est une matrice carrée $\ell \times \ell$, et ceci $\forall\, i$.
Alors nous observons que, si $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_n)$ est une solution du système (1.1), $\hat{x}_i$ est une valeur propre de C_i, et ceci $\forall\, i$. Ensuite il ne nous reste plus qu'à vérifier les différentes combinaisons possibles, comme pour l'algorithme A.

Exemple 15: Considérons de nouveau $\{xy - 3x, y^2 - x, x^2 - 9x\}$ et faisons

$$x \begin{pmatrix} 1 \\ x \\ y \end{pmatrix} \xrightarrow{G} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 9 & 0 \\ 0 & 3 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ x \\ y \end{pmatrix}$$

Les valeurs propres sont 0 et 9. De même, pour

$$y \begin{pmatrix} 1 \\ x \\ y \end{pmatrix} \xrightarrow{G} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 3 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ x \\ y \end{pmatrix}$$

les valeurs propres sont 0 et 3, et, après quelques essais, nous obtenons les solutions du système, i.e. $(0, 0)$ et $(9, 3)$.

1.8 Multiplicité des solutions

Commençons par un exemple et considérons le polynôme à une indéterminée

$$f(x) = (x-1)^5 (x^2+1)^2 (x+5)$$

Ses zéros sont: 1 avec multiplicité 5, i et $-i$ avec multiplicité 2, et -5 comme zéro simple. Si nous comptons toutes les racines de f prises avec leurs multiplicités, nous obtenons 10. Soit de nouveau $E = \{\deg h \mid h \in \langle f \rangle, \ h \not\equiv 0\}$.
Miracle! Nous avons également $\#(\mathbb{N} \setminus E) = 10$.
Bien! Notre prochain but sera de généraliser ce résultat à des systèmes de polynômes à plusieurs variables. En effet, si nous connaissons une base de Gröbner, il est facile de calculer $\#(\mathbb{N}^n \setminus E)$ et donc de savoir combien de solutions il faut obtenir avant de commencer tout calcul.

Rappel: La définition de multiplicité d'une racine d'un polynôme peut être donnée de trois manières équivalentes:

- **Dans l'esprit de l'analyse:** $\hat{x}$ est une racine de multiplicité m d'un polynôme $f \iff \hat{x}$ est solution de tout polynôme dérivé $f^{(\nu)}$, $\nu = 0, \ldots, m-1$, et $f^{(m)}(\hat{x}) \neq 0$.

- **Dans l'esprit de l'analyse numérique:** $\hat{x}$ est une racine de multiplicité m d'un polynôme $f \iff m$ est le nombre maximal de solutions proches de $\hat{x}$ pour le polynôme perturbé $f(x) + \varepsilon = 0$.

- **Dans l'esprit de l'algèbre:** Soit $\hat{x}^{(k)}$ une solution d'un polynôme

$$f(x) = a\prod_{j=1}^{\ell}(x - \hat{x}^{(j)})^{m_j}$$

Nous pouvons écrire

$$\mathfrak{a} = \langle f \rangle = \bigcap_{j=1}^{\ell}\langle (x - \hat{x}^{(j)})^{m_j}\rangle$$

Alors $\hat{x}^{(k)}$ est une racine de multiplicité m_k de $f \iff m = \dim \mathbb{C}[x]/\mathfrak{a}_k$, où $\mathfrak{a}_k = \langle (x - \hat{x}^{(k)})^{m_k}\rangle$.

Dans la suite nous allons considérer $\mathbb{K} = \mathbb{C}$.
Soit le système (1.1). Soient $\mathfrak{a} = \langle f_1, \ldots, f_m\rangle \subset \mathbb{C}[x_1, \ldots, x_n]$ et $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_n)$ une solution de $(\mathcal{F})$.

Opérateurs différentiels: Nous allons noter

$$L = \sum_{\alpha} c_\alpha D^\alpha$$

avec c_α presque tous nuls et $\alpha \in \mathbb{N}^n$, l'opérateur linéaire L, qui s'applique à un polynôme f en étendant linéairement la définition

$$D^\alpha f := \frac{\partial^{\Sigma\alpha_i}}{\partial x_1^{\alpha_1}\cdots\partial x_n^{\alpha_n}} f$$

Notons aussi $D_i = \frac{\partial}{\partial x_i}$ et, par conséquent, $D^\alpha = D_1^{\alpha_1}\cdots D_n^{\alpha_n}$.

Exemple 16: Nous écrivons

$$\frac{\partial^2}{\partial x_1 \partial x_2} f + 5\frac{\partial}{\partial x_2} f + f$$

par Lf, où

$$L := D^{(1,1)} + 5D^{(0,1)} + D^{(0,0)}$$

Définition 10. *Soit $\hat{x}$ une solution du système (1.1). Soit $\mathfrak{a}$ l'idéal engendré par les éléments de $(\mathcal{F})$. Notons par*

$$\mathcal{L}(\mathfrak{a}, \hat{x}) = \operatorname{span}\{1, L_1, L_2, \ldots\}$$

où $\operatorname{span}\{1, L_1, L_2, \ldots\} = \{L \mid \ (Lf)(\hat{x}) = 0 \ \forall\ f \in \mathfrak{a}$ *et L combinaison $\mathbb{C}$-linéaire des L_i*$\}$.

Remarques:

1. $\mathcal{L}(\mathfrak{a}, \hat{x})$ est un $\mathbb{C}$-espace vectoriel.
2. $\hat{x}$ est solution de $(\mathcal{F}) \iff f(\hat{x}) = 0, \forall\ f \in \mathfrak{a} \iff 1 \in \mathcal{L}(\mathfrak{a}, \hat{x})$.
3. Soit $\mathfrak{a} = \langle f \rangle$, où f est un polynôme à une variable. Supposons $f(x) = (x - \hat{x})^m g(x)$ avec $g(\hat{x}) \neq 0$. Alors $\mathcal{L}(\mathfrak{a}, \hat{x}) = \operatorname{span}\{1, D, D^2, \ldots, D^{m-1}\}$.
4. Supposons $f \in \mathfrak{a} \Rightarrow x_1 f(x) \in \mathfrak{a}$. Soit L un opérateur quelconque. Calculons formellement

$$\begin{aligned} L(x_1 f) &= \sum_\alpha c_\alpha D^\alpha (x_1 f) \\ &\overset{Leibniz}{=} \sum_\alpha c_\alpha (x_1 D^\alpha f + 1 \cdot \alpha_1 D^{\alpha - (1,0,\ldots,0)} f) \\ &= x_1 L f + (\partial_1 L) f \end{aligned}$$

où $(\partial_1 L) f = \sum c_\alpha \alpha_1 D^{\alpha-(1,0,\ldots,0)} f$.
Il s'ensuit que $L \in \mathcal{L}(\mathfrak{a}, \hat{x}) \Rightarrow \partial_i L \in \mathcal{L}(\mathfrak{a}, \hat{x})$ et $D_i^k \in \mathcal{L}(\mathfrak{a}, \hat{x}) \Rightarrow D_i^{k-1}, \ldots, D_i^0 \in \mathcal{L}(\mathfrak{a}, \hat{x})$, car $\partial_i D_i^\ell = \ell D_i^{\ell-1}$.

Comme pour les polynômes à une indéterminée, ainsi que ceux à plusieurs variables, il est possible de donner plusieurs définitions de la multiplicité d'une solution. La suivante est tirée de l'analyse et a été donnée par Stetter dans un manuscrit à Vienne.

Définition 11. *Soit $\hat{x}$ solution de $(\mathcal{F})$ et $\mathfrak{a} = \langle f_1, \ldots, f_m \rangle$. La* ***multiplicité*** *de $\hat{x}$ est la dimension de $\mathcal{L}(\mathfrak{a}, \hat{x})$ et se note par $\mu_D(\hat{x})$.*

Exemple 17: Soit $\mathfrak{a} = \langle x^2, y^2 \rangle$. La seule solution est $(\hat{x}, \hat{y}) = (0,0)$. Observons que $\mathcal{L}(\mathfrak{a}, (0,0)) = \operatorname{span}\{1, D_x, D_y, D_{xy}\}$. Donc $\mu_D(0,0) = \#(\mathbb{N}^2 \setminus E) = 4$, où $E = \{\deg f \mid f \in \mathfrak{a},\ f \not\equiv 0\}$.

Exemple 18: Soit $\mathfrak{a} = \langle x^2, y - x \rangle$. $(0,0)$ est une solution. $\mathcal{L}(\mathfrak{a}, (0,0)) = \operatorname{span}\{1, D_x + D_y\}$. Donc $\mu_D(0,0) = 2 = \#(\mathbb{N}^2 \setminus E)$, où $E = \{\deg f \mid f \in \mathfrak{a},\ f \not\equiv 0\}$. Ceci signifie que si nous calculons la dérivée directionnelle en $(0,0)$ de direction $(1,1)$ d'un polynôme dans $\mathfrak{a}$, nous obtenons 0.

Attirons maintenant notre attention sur la **décomposition en fractions premières**.

Lemme 7. *Supposons que le système (1.1) possède seulement un nombre fini de solutions* $\hat{x}^{(1)}, \ldots, \hat{x}^{(\ell)}$. *Posons* $\mathfrak{a}_j = \langle f_1, \ldots, f_m, (x_1 - \hat{x}_1^{(j)})^k, \ldots, (x_n - \hat{x}_n^{(j)})^k \rangle$. *Alors, si* k *est suffisamment grand, nous avons* $\mathfrak{a} = \mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_\ell$, *où* $\mathfrak{a}$ *est l'idéal engendré par les éléments de* $(\mathcal{F})$.

Démonstration. Comme $\forall\ j$ nous avons $\mathfrak{a}_j \supset \mathfrak{a}$, alors $\mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_\ell \supset \mathfrak{a}$. Ce qui prouve une inclusion.
Pour l'autre sens de l'inclusion, soient $\gamma = (\gamma_1, \ldots, \gamma_\ell)$, avec $\gamma_i \in \{1, \ldots, n\}$, et

$$p_\gamma(x) := \prod_{i=1}^{\ell} (x_{\gamma_i} - \hat{x}_{\gamma_i}^{(i)})$$

p_γ s'annule sur toute solution de $(\mathcal{F})$. Donc, par le théorème 4, $\exists\ k_\gamma$ tel que $p_\gamma^{k_\gamma} \in \mathfrak{a}$. Soit $k = \max k_\gamma$. Alors $(p_\gamma)^k \in \mathfrak{a}\ \forall\ \gamma$.
Soit $f \in \mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_\ell$. Alors nous pouvons écrire

$$f(x) = \sum_\gamma u_\gamma(x) p_\gamma(x)^k + g(x)$$

avec $g \in \mathfrak{a}$. Donc $f \in \mathfrak{a}$ si k est suffisamment grand. $\square$

Une autre définition possible pour la multiplicité d'une solution est celle d'inspiration algébrique.

Définition 12. *Soit* $\mathfrak{a} = \mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_\ell$ *la décomposition du lemme 7. La valeur*

$$\mu_A(\hat{x}^{(j)}) = \dim \mathbb{C}[x]/\mathfrak{a}_j$$

est la ***multiplicité*** *de* $\hat{x}^{(j)}$.

Remarquons qu'à partir d'un certain valeur de k, l'idéal $\mathfrak{a}_j$ ne change plus. Donc μ_A est bien définie. Pour calculer cette dimension, nous pouvons procéder d'abord par la détermination d'une base de Gröbner pour $\mathfrak{a}_j$. Puis nous calculons $E_j = \{\deg f \mid f \in \mathfrak{a}_j\}$ et, enfin, nous obtenons

$$\mu_A(\hat{x}^{(j)}) = \dim \mathbb{C}[x]/\mathfrak{a}_j = \#(\mathbb{N}^n \setminus E_j)$$

La chose surprenante, affirmée par Stetter dans son manuscrit, est que les deux définitions sont équivalentes. Ce résultat est tellement surprenant, qu'il mérite d'être énoncé dans un théorème. Par suite, nous pouvons définir la multiplicité d'une solution $\hat{x}$ de manière essentiellement unique et la noter tout simplement par $\mu(\hat{x})$.

Théorème 14. *Les deux définitions de multiplicité sont équivalentes, i.e.* $\mu_A = \mu_D$.

Démonstration. Soit $(\mathcal{F})$ le système (1.1). Soient $\hat{x}^{(j)}$ une solution de $(\mathcal{F})$ et $\mathfrak{a} = \langle f_1, \ldots, f_m \rangle$. Comme dans la décomposition du lemme 7, posons

$$\mathfrak{a}_j = \langle f_1, \ldots, f_m, (x_1 - \hat{x}_1^{(j)})^k, \ldots, (x_n - \hat{x}_n^{(j)})^k \rangle$$

et montrons que $\mathcal{L}(\mathfrak{a}, \hat{x}^{(j)}) = \mathcal{L}(\mathfrak{a}_j, \hat{x}^{(j)})$. Ensuite nous allons voir une méthode pour construire $\mathcal{L}(\mathfrak{a}_j, \hat{x}^{(j)})$, pour avoir finalement que $\mathbb{C}[X] \backsimeq \mathfrak{a}_j \oplus \mathfrak{a}_j^\perp$, où $\mathfrak{a}_j^\perp$ est l'ensemble des polynômes orthogonaux à $\mathfrak{a}_j$ selon une forme bilinéaire que nous aurons définie. Après avoir réalisé ceci nous aurons terminé.

1. $\boldsymbol{\mathcal{L}(\mathfrak{a}, \hat{x}^{(j)}) = \mathcal{L}(\mathfrak{a}_j, \hat{x}^{(j)})}$
   $\mathfrak{a}_j \supset \mathfrak{a} \Rightarrow \mathcal{L}(\mathfrak{a}, \hat{x}^{(j)}) \supset \mathcal{L}(\mathfrak{a}_j, \hat{x}^{(j)})$, ce qui prouve une inclusion.
   Pour l'autre sens de l'inclusion, soit $L \in \mathcal{L}(\mathfrak{a}, \hat{x}^{(j)})$.
   Donc $Lf(\hat{x}^{(j)}) = 0 \ \forall\ f \in \mathfrak{a}$, en particulier $(L(x_i - \hat{x}_i^{(j)})^k)(\hat{x}^{(j)}) = 0$, si k est suffisamment grand.

2. **Isomorphisme entre** $\boldsymbol{\mathcal{L}(\mathfrak{a}_j, \hat{x}^{(j)})}$ **et** $\boldsymbol{\mathfrak{a}_j{}^\perp}$
   Remplaçons dans la notation $\mathfrak{a}_j$ par $\mathfrak{a}$ et supposons que $\mathfrak{a}$ possède $\hat{x} = 0$ comme unique solution. Nous pouvons le faire par la translation $x \rightsquigarrow x - \hat{x}$. Soit

$$f(x) = \sum_\alpha a_\alpha x^\alpha$$

   un polynôme à coefficients complexes. Posons

$$L_f = \sum_\alpha \frac{1}{\alpha!} \cdot \overline{a_\alpha} \cdot D^\alpha$$

   où $\alpha! = \alpha_1! \cdots \alpha_n!$.
   Calculons pour un polynôme $g(x) = \sum b_\beta x^\beta$

$$\begin{aligned}(L_f g)(0) &= \sum_\alpha \overline{a_\alpha} \sum_\beta b_\beta \cdot \frac{1}{\alpha!} \cdot \underbrace{(D^\alpha x^\beta)(0)}_{\neq 0 \iff \alpha=\beta} \\ &= \sum_\alpha \overline{a_\alpha} \cdot b_\alpha\end{aligned}$$

   Remarquons que ceci est équivalent au produit hermitien standard sur $\mathbb{C}[x]$ de g avec f. Nous avons donc que

$$\begin{aligned}\mathcal{L}(\mathfrak{a}, 0) &\overset{déf}{=} \{L \mid \ (Lg)(0) = 0 \ \forall\ g \in \mathfrak{a}\} \\ &\backsimeq \{f \in \mathbb{C}[x] \mid < g|f >= 0 \ \forall\ g \in \mathfrak{a}\} \\ &= \mathfrak{a}^\perp\end{aligned}$$

3. $\mathbb{C}[\boldsymbol{x}] \backsimeq \mathfrak{a}_j \oplus \mathfrak{a}_j{}^\perp$
   Egalement dans cette partie de la preuve, nous remplaçons dans la notation $\mathfrak{a}_j$ par $\mathfrak{a}$ et nous supposons que $\mathfrak{a}$ possède $\hat{x} = 0$ comme unique solution.
   A voir: $\forall\ f \in \mathbb{C}[x]\ \exists!\ g \in \mathfrak{a}$ et $h \in \mathfrak{a}^\perp$ tels que $f = g + h$.

Dans ce cas nous aurions ${}^{\mathbb{C}[x]}\!\big/_{\mathfrak{a}} \backsimeq \mathfrak{a}^{\perp}$, $[f] = f + \mathfrak{a} \leftrightarrow h$.
Pour réaliser ceci, nous allons construire une base orthogonale de $\mathfrak{a}$ et ensuite la compléter en une base de $\mathbb{C}[x]$.
Par le *Nullstellensatz de Hilbert* (théorème 4), comme les polynômes x_i s'annulent en 0, il existe des $k_i \in \mathbb{N}$ tels que $x_i^{k_i} \in \mathfrak{a}$. Soit $k = \max\{k_1, \ldots, k_n\}$. Alors $x_i^k \in \mathfrak{a}$, $i = 1, \ldots, n$.
Considérons les polynômes $g_\alpha(x) = x^\alpha$. Nous avons que $g_\alpha \in \mathfrak{a}$ si $\exists\ \alpha_i \geq k$. De plus, les g_α sont orthonormaux, car

$$< x^\alpha | x^{\alpha'} > \overset{si\ \alpha \neq \alpha'}{=} 1 \cdot 0 + 0 \cdot 1 = 0$$

Nous avons trouvé une base orthogonale pour la partie

$$E_\infty = \{\alpha = \deg f \mid\ f \in \mathfrak{a},\ \exists\ \alpha_i \geq k\}$$

(c'est la région hachurée finement sur la figure 3).
Cherchons la base de la partie $E \setminus E_\infty$, où $E = \{\deg f \mid\ f \in \mathfrak{a},\ f \not\equiv 0\}$ (dans la figure 3, il s'agit de la partie hachurée par une large grille).
Pour chaque $\alpha \in E \setminus E_\infty$, soient $\widehat{g_\alpha} \in \mathfrak{a}$ tels que $\deg \widehat{g_\alpha} = \alpha$.
De là, nous utilisons la méthode de Grahm-Schmidt pour orthogonaliser les $\widehat{g_\alpha}$ à partir des $\widehat{g_\alpha}$ avec les α géométriquement les plus proches de $0 \in \mathbb{N}^n$. L'orthogonalisation se fait par rapport aux autres $\widehat{g_\alpha}$ et aux x^β, $\beta \in E_\infty$. Nous obtenons $\{g_\alpha \mid\ \alpha \in E\}$, qui est une base orthonormale de $\mathfrak{a}$. Regardons

$$\{g_\alpha \mid\ \alpha \in E\} \sqcup \{x^\alpha \mid\ \alpha \in \mathbb{N}^n \setminus E\}$$

qui est une base de $\mathbb{C}[x]$.
Pour $\alpha \in \mathbb{N}^n \setminus E$, nous orthogonalisons x^α par rapport à $\mathfrak{a}$.

$$x^\alpha - \sum_{\beta \in E} c_\beta = g_\beta(x) \quad \in \mathbb{C}[x] \tag{1.15}$$

où $c_\beta =< x^\alpha | g_\beta > \in \mathbb{C}$. Comme $\mathbb{N}^n \setminus E$ est un ensemble fini par hypothèse ("... et supposons que $\mathfrak{a}$ possède $\hat{x} = 0$ comme unique solution"), seulement un nombre fini de c_β est différent de 0. Il s'ensuit que la somme dans (1.15) est bien définie.

Maintenant nous pouvons écrire en toute sécurité

$$\begin{aligned}
\mu_A(\hat{x}^{(j)}) &\overset{déf}{=} \dim {}^{\mathbb{C}[x]}\!\big/_{\mathfrak{a}_j} \\
&\overset{3)}{=} \dim \mathfrak{a}_j^{\perp} \\
&\overset{2)}{=} \dim \mathcal{L}(\mathfrak{a}_j, \hat{x}^{(j)}) \\
&\overset{1)}{=} \dim \mathcal{L}(\mathfrak{a}, \hat{x}^{(j)}) \\
&\overset{déf}{=} \mu_D(\hat{x}^{(j)})
\end{aligned}$$

Ce qui termine la preuve. $\square$

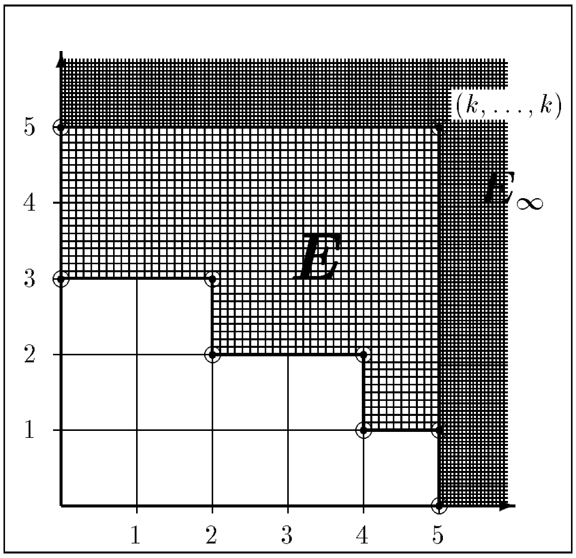

FIGURE 3. *Exemple graphique d'ensembles E et E_∞.*

Le théorème qui suit nous montre un résultat déjà conjecturé au début de ce paragraphe et illustré avec plusieurs exemples.

Théorème 15. *Soit $\mathfrak{a} = \langle f_1, \ldots, f_m \rangle$ un idéal avec un nombre fini de solutions distinctes $\hat{x}^{(1)}, \ldots, \hat{x}^{(t)}$. Fixons un terme ordre. Alors*

$$\#(\mathbb{N}^n \setminus E) = \sum_{i=1}^{t} \mu(\hat{x}^{(i)})$$

où $E = \{\deg f \mid \ f \in \mathfrak{a},\ f \not\equiv 0\}$.

Démonstration. Nous allons montrer que

$$\mathbb{C}[x]/\mathfrak{a} \cong \mathbb{C}[x]/\mathfrak{a}_1 \times \cdots \times \mathbb{C}[x]/\mathfrak{a}_t$$

où $\mathfrak{a} = \mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_t$ est la décomposition du lemme 7. Ceci va impliquer par un résultat élémentaire d'algèbre linéaire que

$$\begin{aligned}
\#(\mathbb{N}^n \setminus E) &= \dim \mathbb{C}[x]/\mathfrak{a} \\
&= \sum_{i=1}^{t} \dim \mathbb{C}[x]/\mathfrak{a}_i \\
&= \sum_{i=1}^{t} \mu(\hat{x}^{(i)})
\end{aligned}$$

avec $E = \{\deg f \in \mathfrak{a} \mid \ f \not\equiv 0\}$.
Considérons l'application

$$\begin{array}{rcl} \phi \ : \ \mathbb{C}[x] & \longrightarrow & \mathbb{C}[x]/_{\mathfrak{a}_1} \times \cdots \times \mathbb{C}[x]/_{\mathfrak{a}_t} \\ f & \longmapsto & (f + \mathfrak{a}_1, \ldots, f + \mathfrak{a}_t) \end{array}$$

et calculons son noyau:

$$\begin{array}{rcl} \ker \phi & = & \{f \in \mathbb{C}[x] \mid \ f + \mathfrak{a}_i = \mathfrak{a}_i \ \forall \ 1 \leq i \leq t\} \\ & = & \{f \in \mathbb{C}[x] \mid \ f \in \mathfrak{a}_i \ \forall \ 1 \leq i \leq t\} \\ & = & \bigcap_{i=1}^{t} \{f \in \mathbb{C}[x] \mid \ f \in \mathfrak{a}_i\} \\ & = & \bigcap_{i=1}^{t} \mathfrak{a}_i \\ & = & \mathfrak{a} \end{array}$$

Donc l'application

$$\begin{array}{rcl} \overline{\phi} \ : \quad \mathbb{C}[x]/_{\mathfrak{a}} & \longrightarrow & \mathbb{C}[x]/_{\mathfrak{a}_1} \times \cdots \times \mathbb{C}[x]/_{\mathfrak{a}_t} \\ f + \mathfrak{a} = [f] & \longmapsto & (f + \mathfrak{a}_1, \ldots, f + \mathfrak{a}_t) \end{array}$$

est bien définie et est injective. Voyons, de plus, que $\overline{\phi}$ est un isomorphisme.
Pour la suite, nous supposons de travailler avec des polynômes à une variable, afin de ne pas nous encombrer avec les notations. Pour le cas plus en général, on adapte.
Soient

$$\ell_i(x) = \prod_{j \neq i} \frac{(x - \hat{x}^{(j)})}{(\hat{x}^{(i)} - \hat{x}^{(j)})} \ \in \bigcap_{j \neq i} \mathfrak{a}_j$$

les polynômes d'interpolation de Lagrange. Les seules racines de ces polynômes sont les solutions $\hat{x}^{(j)}$ avec $j \neq i$, et ceci pour chaque ℓ_i. Nous avons aussi que le système $\{\ell_i(x) = 0\}_{i=1}^{t}$ n'admet pas de solutions. Il est donc possible, par le théorème 3, de trouver une décomposition

$$1 = \sum_{i=1}^{t} p_i(x) \ell_i(x)$$

Soit

$$([f_1], \ldots, [f_t]) \in \mathbb{C}[x]/_{\mathfrak{a}_1} \times \cdots \times \mathbb{C}[x]/_{\mathfrak{a}_t}$$

arbitraire. Il faut trouver $f \in \mathbb{C}[x]$ tel que $f + \mathfrak{a}_i = f_i + \mathfrak{a}_i$, i.e. $f - f_i \in \mathfrak{a}_i$, $\forall \ 1 \leq i \leq t$.
L'idée est de considérer

$$f(x) := \sum_{i=1}^{t} p_i(x) \ell_i(x) f_i(x)$$

Vérifions que $f - f_1 \in \mathfrak{a}_1$. La preuve pour les autres est analogue.

$$\begin{aligned} f - f_1 &= \sum_{i=1}^{t} p_i \ell_i f_i \; - \; f_1 \cdot 1 \\ &= \sum_{i=1}^{t} p_i \ell_i f_i \; - \; f_1 \cdot \sum_{i=1}^{t} p_i \ell_i \\ &= \sum_{i=2}^{t} \ell_i \cdot (p_i f_i - p_i f_1) \end{aligned}$$

est un élément de $\mathfrak{a}_1$, puisque les ℓ_i avec $i \neq 1$ le sont. Ce qui prouve que $\overline{\phi}$ est aussi surjective.

Enfin, par construction, il est évident que $\overline{\phi}$ est un homomorphisme de $\mathbb{C}$-modules. □

1.9 Systèmes polynomiaux perturbés

Le sujet traité dans cette partie du cours consiste à voir si les méthodes présentées jusqu'à maintenant sont efficaces pour résoudre des problèmes pratiques, pour lesquels, nous n'avons souvent pas de certitude à propos de l'exactitude des données fournies. Par exemple, pour un certain problème, les données que l'on nous livre sont des mesures qui, en général, se révèlent approximatives pour des raisons pratiques.

Considérons le système (1.1). Nous pouvons voir le problème de la recherche des solutions de ce système, comme la recherche des zéros de l'application

$$\begin{array}{cccc} f : & \mathbb{K}^n & \longrightarrow & \mathbb{K}^m \\ & \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} & \longmapsto & \begin{pmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{pmatrix} \end{array} \tag{1.16}$$

Sauf d'autres indications, nous allons noter $\mathfrak{a} = \langle f_1, \ldots, f_m \rangle$ et par $\hat{x}$ une possible solution de f.

Soit ε petit. Nous allons étudier le système perturbé

$$f(x) + \varepsilon \cdot g(x, \varepsilon) = 0$$

que nous allons encore noter par $f(x, \varepsilon)$.

Exemple 19: Soient $f(x) = x^3$ un polynôme à une variable et $g(x, \varepsilon) = x^4 + \varepsilon x + 1$. La seule solution de $f(x) = 0$ est $\hat{x} = 0$, avec multiplicité 3. Si nous perturbons f

$$f(x) + \varepsilon g(x, \varepsilon) = x^3 + \varepsilon x^4 + \varepsilon + \varepsilon^2 x = 0$$

nous avons (voir figure 4) trois racines complexes "proches" de l'origine (il s'agit de $-\sqrt[3]{\varepsilon}$ et de $\sqrt[3]{\varepsilon}(1/2 \pm i\sqrt{3}/2)$) et une quatrième beaucoup plus loin $(-1/\varepsilon)$.

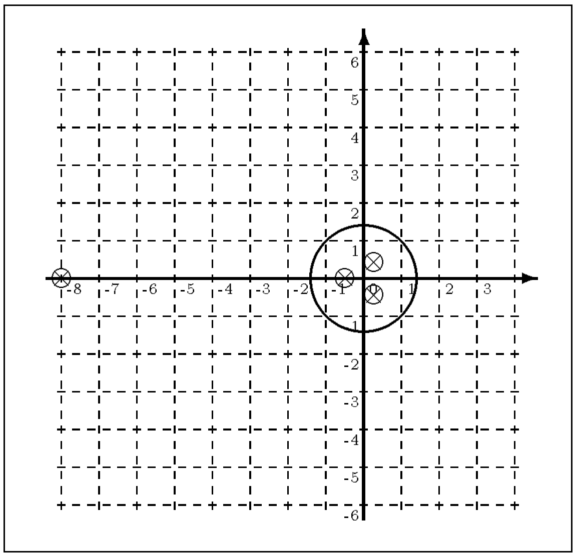

FIGURE 4. *Position des racines de l'exemple 19 avec* $\varepsilon = 1/8$.

1.9.1 Racines simples ($\mu(\hat{x}) = 1$)

Lemme 8. *Soit* $\hat{x}$ *une racine d'un système de polynômes* $f(x) = 0$. *Alors* $\mu(\hat{x}) = 1 \iff \operatorname{rang} f'(\hat{x}) = n$, *i.e.* $f'(\hat{x})$ *possède un rang maximal.*

Démonstration. Soit $\mathfrak{a}$ l'idéal engendré par les composantes de f et supposons par l'absurde que $\mu(\hat{x}) > 1$. Alors

$$\begin{aligned}
\mu(\hat{x}) > 1 \quad & \overset{déf}{\iff} \quad \dim \mathcal{L}(\mathfrak{a}, \hat{x}) > 1 \\
& \iff \quad \exists\ L \in \mathcal{L}(\mathfrak{a}, \hat{x}),\ L \notin \operatorname{span}\{1\},\ \text{t.q. } \forall\ g \in \mathfrak{a} \text{ on a } (Lg)(\hat{x}) = 0 \\
& \iff \quad \exists\ d \in \mathbb{C}^n \setminus \{0\} \text{ t.q. } L = \sum_{i=1}^{n} d_i \frac{\partial}{\partial x_i} \in \mathcal{L}(\mathfrak{a}, \hat{x}) \\
& \overset{(*)}{\iff} \quad \exists\ d \in \mathbb{C}^n \setminus \{0\} \text{ t.q. } f'(\hat{x}) \cdot d = 0
\end{aligned}$$

Montrons l'équivalence $(*)$.

$\Rightarrow$: trivial, car $(f'(\hat{x}) \cdot d)_j = \sum_{i=1}^{n} d_i \frac{\partial f_j}{\partial x_i}(\hat{x}) = ((Lf_j)(\hat{x})) = 0$.

$\Leftarrow$: Soit $g = \sum_{j=1}^{m} u_j f_j \in \mathfrak{a}$.

$$(Lg)(\hat{x}) \quad = \quad \sum_{i=1}^{n} d_i \frac{\partial}{\partial x_i} g(\hat{x})$$

$$
\begin{aligned}
&\overset{Leibniz}{=} \sum_{i=1}^{n}\sum_{j=1}^{m} d_i \frac{\partial u_j}{\partial x_i}\underbrace{f_j(\hat{x})}_{=0} + \sum_{i=1}^{n}\sum_{j=1}^{m} d_i u_j \frac{\partial f_j}{\partial x_i}(\hat{x}) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} d_i u_j \frac{\partial f_j}{\partial x_i}(\hat{x}) \\
&= 0
\end{aligned}
$$

D'où $\mu(\hat{x}) > 1 \iff \operatorname{rang} f'(\hat{x}) < n$. □

Pour une application de ce résultat, considérons $f(x, \varepsilon) = 0$ dans un voisinage de $\hat{x}$ et regardons les deux possibilités suivantes:

1. $\boldsymbol{m > n}$ **:** Il est possible que $f(x, \varepsilon) = 0$ ne possède pas de solutions près de $\hat{x}$. Si le système est surdéterminé il y a peu de chances que $f(x, \varepsilon) = 0$ admette une solution [11]. Nous pouvons espérer, par contre, de trouver des valeurs de x telles que $\|f(x, \varepsilon)\|$ soit minimale. Ce problème sera analysé plus en détail dans le prochain chapitre.

2. $\boldsymbol{m = n}$ **:** Ici nous pouvons appliquer le théorème des fonctions implicites à

$$f(x, \varepsilon) = 0$$

Nous avons $f(\hat{x}, 0) = 0$ et $\frac{\partial}{\partial x}(\hat{x}, 0) = f'(\hat{x})$, qui est inversible. Nous pouvons donc définir une application infiniment différentiable $x(\varepsilon)$ près de $\hat{x}$ telle que $x(0) = \hat{x}$ et $f(x(\varepsilon), \varepsilon) = 0$.

Si nous voulons résumer les résultats ci-dessus nous pouvons dire que si $\hat{x}$ est une racine simple de $f(x) = 0$, alors le système perturbé $f(x, \varepsilon) = 0$ possède au plus une solution proche de $\hat{x}$.

1.9.2 Racines multiples ($\boldsymbol{\mu(\hat{x}) > 1}$)

Les affirmations suivantes sont données sans démonstration. Par contre, elles seront illustrées par des exemples.

Si $\hat{x}$ est une racine de multiplicité $\mu > 1$ de $f(x)$, alors:

- Le système perturbé possède au plus μ racines (comptées avec leurs multiplicités) proches de $\hat{x}$.
- Il existe une perturbation pour laquelle le système perturbé possède exactement μ racines simples proches de $\hat{x}$.

11. Si je vous propose un système et que je le perturbe

$$\left\{\begin{array}{l} x-1=0 \\ x^2-1=0 \end{array}\right. \quad \rightsquigarrow \quad \left\{\begin{array}{l} x-1=\varepsilon_1 \\ x^2-1=\varepsilon_2 \end{array}\right.$$

alors vous pouvez trouver une solution au système perturbé si et seulement si $(1+\varepsilon_1)^2 - 1 = \varepsilon_2$. Ce qui ne sera pas le cas. Ha, ha, ha!

Exemple 20: Soit $\{f_1, \ldots, f_m\}$ une base de Gröbner. Fixons un terme ordre. Soit $E = \{\deg f \mid \ f \in \mathfrak{a},\ f \not\equiv 0\}$. Regardons le système perturbé [12]

$$(\mathcal{F}_\varepsilon) \begin{cases} f_1(x_1, \ldots, x_n) + \varepsilon \cdot \sum\limits_{\alpha \in \mathbb{N}^n \setminus E} c_{1\alpha}(\varepsilon)\ x^\alpha = 0 \\ \quad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ f_m(x_1, \ldots, x_n) + \varepsilon \cdot \sum\limits_{\alpha \in \mathbb{N}^n \setminus E} c_{m\alpha}(\varepsilon)\ x^\alpha = 0 \end{cases} \tag{1.17}$$

Maintenant changeons de terme ordre. Laissons invarié l'ordre sur E et modifions-le sur $\mathbb{N}^n \setminus E$. La seule chose à laquelle il faut faire attention, pour que le nouvel ordre reste un terme ordre, est que l'indice augmente à chaque fois que l'on s'éloigne de l'origine (voir figure 5).

Soit $E(\varepsilon) = \{\deg g \mid \ g \in \langle f(x, \varepsilon)\rangle,\ g \not\equiv 0\}$. Nous avons $E(\varepsilon) \supseteq E$, donc aussi $\mathbb{N}^n \setminus E(\varepsilon) \subseteq \mathbb{N}^n \setminus E$. Donc, si nous perturbions un système, nous ne pourrions pas augmenter le nombre de solutions.

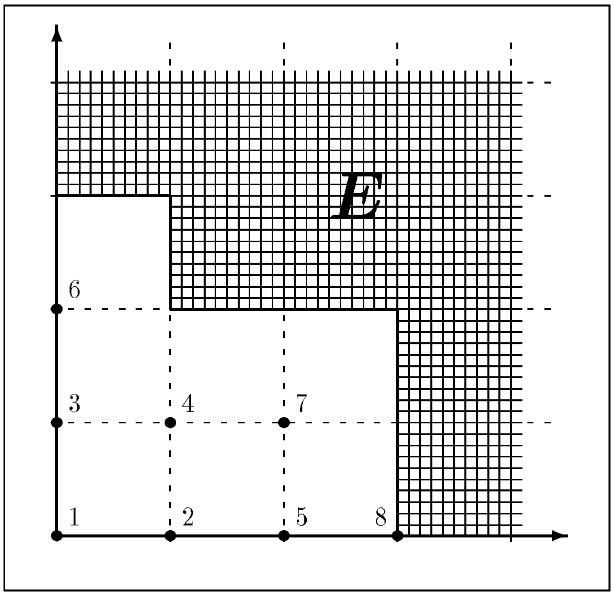

FIGURE 5.

Exemple 21: Considérons un cas qui a une seule solution, que nous supposons être nulle. Prenons, par exemple,

$$\begin{cases} x^6 & = & 0 \\ x^2y^2 & = & 0 \\ y^4 & = & 0 \end{cases}$$

12. Il nous suffit de prendre des valeurs pour α dans $\mathbb{N}^n \setminus E$, puisque nous pouvons supposer, après translation, que $\hat{x} = 0$. Dans ce cas, si $\alpha \in E$, alors α est “grand” et les termes en x^α deviennent négligeables dans un petit voisinage de 0.

Nous avons que $\mu_D(0) = 16$. Cherchons une perturbation telle que $(\varepsilon\hat{x}_i, \varepsilon\hat{y}_j)$ soient des racines distinctes pour $(i,j) \in \mathbb{N}^2 \setminus E$. Ceci est donné par

$$\begin{cases} x^6 + \varepsilon g_1(x,\varepsilon) = (x - \varepsilon\hat{x}_1)\cdots(x - \varepsilon\hat{x}_6) \\ x^2y^2 + \varepsilon g_2(x,y,\varepsilon) = (x - \varepsilon\hat{x}_1)(x - \varepsilon\hat{x}_2)(y - \varepsilon\hat{y}_1)(y - \varepsilon\hat{y}_2) \\ y^4 + \varepsilon g_3(y,\varepsilon) = (y - \varepsilon\hat{y}_1)\cdots(y - \varepsilon\hat{y}_4) \end{cases}$$

1.10 Amélioration de l'algorithme de Buchberger

Avant de nous lancer dans des papotages sans fin, commençons par rappeler que l'algorithme de Buchberger pour la construction d'une base de Gröbner repose sur la constatation suivante:

$$G \text{ est une base de Gröbner } \iff \forall\, g_i, g_j \in G \;:\; S(g_i, g_j) \xrightarrow{G} 0$$

(Voir théorème 10 à la page 34).

1.10.1 Amélioration 1

Une première bonne idée est d'appliquer l'algorithme de réduction d'une base à chaque fois que l'on ajoute un polynôme à la base.

1.10.2 Amélioration 2

Cette deuxième amélioration consiste à reformuler le théorème 10, afin de pouvoir l'utiliser d'une manière plus efficace.
Considérons le système (1.1). Soit, comme d'habitude, $G = \{f_1, \ldots, f_m\}$. Alors, pour vérifier que G est une base de Gröbner, il suffit de tester $S(f,g) \xrightarrow{G} 0$ seulement pour les couples de polynômes qui sont appelés G-voisins.

Définition 13. *Soient f, $g \in G$, où G est une famille de polynômes. Alors f et g sont **G-voisins** $\iff \nexists\, h \in G$ tel que* $\text{ppcm}(tp(f), tp(g)) = x^\alpha tp(h)$, *avec $\alpha \geq 0$.*

Pour mieux comprendre ce concept, regardons de nouveau la figure 2 à la page 33.

Exemple 22: Soit $G = \{f_1, f_2, f_4\}$. Alors f_1 et f_2 sont G-voisins, de même pour f_2 et f_4. Cependant f_1 et f_4 ne le sont pas car $\deg \text{ppcm}(f_1, f_4) = \deg x_1x_2^2 f_2$.

D'autre part, que nous pouvons comprendre avec cet exemple à deux indéterminées, que, en général, pour une famille de m polynômes, il faut calculer $m(m-1)/2$ polynômes de soustraction. Mais, avec cette amélioration, il suffit de faire le calcul pour $m-1$ couples.

Vérifions maintenant que tout ceci est légal.

Théorème 16. *Soit* $G = \{f_1, \ldots, f_m\}$ *un générateur d'un idéal.*
Alors G *est une base de Gröbner* $\Longleftrightarrow$ $\forall\ f, g$ G*-voisins on a* $S(f,g) \xrightarrow{G} 0$.

Démonstration. La nécessité est évidente et suit directement par la définition de base de Gröbner. Pour la suffisance, nous donnons seulement l'idée, car la démonstration est similaire à celle du théorème 10.
Soit donc $h \in \langle G \rangle$. A voir $h \xrightarrow{G} 0$.
Comme dans la preuve du théorème 10, supposons h réduit et, par l'absurde, $h \not\equiv 0$.

$$h(x) = \sum_{\substack{i,\alpha \\ finie}} c_{i\alpha} x^\alpha f_i(x)$$

Soient aussi δ et $\ell(x)$ définis comme dans la preuve du théorème 10. Nous avons donc que $\ell(x)$ est sous la forme (1.13). Supposons maintenant que dans cette écriture se trouve un couple de f_i et f_j qui ne sont pas G-voisins. Alors $\exists\ f_t \in G$ tel que $\mathrm{ppcm}(tp(f_i), tp(f_j)) = x^{\xi_{ij}} tp(f_t) = x^{\zeta_{ij}}$, pour des certains ξ_{ij} et ζ_{ij}. Nous pouvons supposer que f_t et f_i soient G-voisins. Nous avons donc que

$$\begin{aligned}
S(f_i, f_j) &= x^{\zeta_{ij} - \deg f_i} f_i(x) - x^{\zeta_{ij} - \deg f_j} f_j(x) \\
&= x^{\zeta_{ij} - \deg f_i} f_i(x) - x^{\zeta_{ij} - \deg f_t} f_t(x) + \\
&\qquad + x^{\zeta_{ij} - \deg f_t} f_t(x) - x^{\zeta_{ij} - \deg f_j} f_j(x) \\
&= x^{\nu_{it}} S(f_i, f_t) + x^{\nu_{tj}} S(f_t, f_j)
\end{aligned}$$

En répétant cet argument, nous pouvons supposer que la somme de (1.13) se fait uniquement sur des couples de polynômes G-voisins, vu que G ne contient qu'un nombre fini d'éléments. Puis nous continuons la preuve exactement de la même manière que celle du théorème 10. $\square$

1.10.3 Amélioration 3

Ici nous allons voir quand il est inutile de calculer un polynôme de soustraction.

Proposition 1. *Soient* f *et* $g \in \mathbb{K}[x]$*. Alors, si* $tp(f)$ *et* $tp(g)$ *sont premiers entre eux,* $S(f,g) \xrightarrow{\{f,g\}} 0$.

Démonstration. Ecrivons

$$f(x) = tp(f) + \hat{f}(x) \qquad\qquad g(x) = tp(g) + \hat{g}(x)$$

Par suite, nous avons

$$\begin{aligned}
S(f,g) &= tp(g) \cdot f - tp(f) \cdot g \\
&= tp(g) \cdot \hat{f} - tp(f) \cdot \hat{g} \\
&\xrightarrow{\{f,g\}} -\hat{g}\hat{f} + \hat{f}\hat{g} = 0
\end{aligned}$$

$\square$

1.10.4 Amélioration 4

Pour le système (1.1) nous pouvons essayer de factoriser des f_i.
Par exemple, si $f_1(x) = \hat{f}_1(x)\hat{\hat{f}}_1(x)$, alors

$$\left\{\begin{array}{c} f_1(x) = 0 \\ \vdots \\ f_m(x) = 0 \end{array}\right. \iff \left\{\begin{array}{c} \hat{f}_1(x) = 0 \\ f_2(x) = 0 \\ \vdots \\ f_m(x) = 0 \end{array}\right. \text{ou} \left\{\begin{array}{c} \hat{\hat{f}}_1(x) = 0 \\ f_2(x) = 0 \\ \vdots \\ f_m(x) = 0 \end{array}\right.$$

Chercher des solutions pour les deux systèmes à droite est en général plus simple que résoudre directement (1.1), car $\deg \hat{f}_1$ et $\deg \hat{\hat{f}}_1$ sont plus petits de $\deg f_1$.

1.11 Racines de polynômes à une variable

Dans ce dernier paragraphe du chapitre, nous allons porter notre intérêt sur des méthodes qui nous aideront à trouver toutes les racines $\hat{x}$ d'un polynôme $f(x)$ à une indéterminée. Comme première idée, il y a la méthode de Newton[13]. Mais le problème de cet algorithme est que pour des racines de multiplicité élevée la convergence est lente. Toutefois nous pouvons contourner ce handicap de la manière suivante. Calculons

$$g(x) := \frac{f(x)}{\mathrm{pgcd}(f, f')}$$

Maintenant nous obtenons que f et g ont les même racines, mais par contre que toutes les racines de g sont simples!

Exemple 23:

$$\begin{aligned} f(x) &= (x-1)^3(x-2)(x+1)^5 \\ f'(x) &= (x-1)^2(x+1)^4(9x^2-18x+3) \\ \mathrm{pgcd}(f, f') &= (x-1)^2(x+1)^4 \\ g(x) &= (x-1)(x-2)(x+1) \end{aligned}$$

Donc à partir de maintenant nous pouvons supposer que toutes les racines de $f(x)$ sont simples.

13. L'algorithme de Newton pour la recherche des racines d'un polynôme $f(x)$ consiste à se donner une valeur de départ x_0 arbitraire et de construire une suite $\{x_k\}_k$, où

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

et à remarquer que la suite converge vers une racine de $f(x)$, s'il y en a une.

1.11.1 Estimation des racines

Ici, nous voudrions chercher une valeur R telle que toutes les racines $\hat{x}$ de $f(x)$ vérifient $|\hat{x}| \leq R$. Puis, une fois avoir fait ceci, nous pourrions trouver les solutions avec une méthode connue, comme par exemple la bissection...
L'idée pour trouver un tel R est de transformer le problème en un problème de recherche de valeurs propres.
Soit

$$f(x) = a_n x^n + \cdots + a_0$$

où $a_n \neq 0$. Alors

$$\begin{aligned} f(x) = 0 \quad & \Longleftrightarrow \quad x^n = -\frac{a_{n-1}}{a_n} x^{n-1} - \cdots - \frac{a_0}{a_n} \\ & \Longleftrightarrow \quad x \begin{pmatrix} x^{n-1} \\ x^{n-2} \\ \vdots \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} -\frac{a_{n-1}}{a_n} & -\frac{a_{n-2}}{a_n} & \cdots & -\frac{a_0}{a_n} \\ 1 & 0 & \cdots & 0 \\ & \ddots & & \vdots \\ 0 & & 1 & 0 \end{pmatrix}}_{:=A} \begin{pmatrix} x^{n-1} \\ \vdots \\ 1 \end{pmatrix} \end{aligned}$$

D'où $\hat{x}$ est une racine de f si et seulement si $\hat{x}$ est une valeur propre de A.
A partir de là, pour l'estimation, nous pouvons utiliser un théorème bien connu, dont la preuve est laissée comme exercice ou, pour ceux qui ne veulent pas trop se fatiguer, comme une attentive recherche dans le cours d'introduction à l'analyse numérique, dont le polycopié peut être retrouvé dans les pages WEB de Ernst Hairer[14].

Théorème 17. *(Théorème de Gershgorine)*
Soit $A = (a_{ij})$ une matrice carrée d'ordre finie. Soit $\lambda \in \operatorname{spec} A$. Alors $\exists$ i tel que $|\lambda - a_{ii}| \leq \sum_{i \neq j} |a_{ij}|$. □

Donc, si $\hat{x}$ est une racine, alors

$$|\hat{x}| \leq \max\left\{1, \sum_{i=0}^{n-1} \left|\frac{a_i}{a_n}\right|\right\}$$

et si nous appliquons le théorème de Gershgorine à A^T, nous obtenons

$$|\hat{x}| \leq \max_{i=1,\ldots,n-1} \left(1 + \left|\frac{a_i}{a_n}\right|\right)$$

Nous pouvons prendre pour R la plus petite des deux valeurs.

Voilà! Maintenant que nous avons trouvé R, pouvons-nous nous dire satisfaits? Oui et non. Pourquoi? Parce que d'un côté ces deux estimations ne sont pas optimales, mais

14. http://www.unige.ch/math/hairer

d'un autre côté elles sont faciles à calculer.

Exemple 24: Soit $f(x) = x^n - 2^n$ pour $n \in \mathbb{N}$. Bien évidemment les racines de f satisfont $|\hat{x}| = 2$. Toutefois, les deux éstimations ci-dessus donnent $|\hat{x}| \leq \max\{1, 2^n\} = 2^n$ comme meilleur résultat.
Pour améliorer ceci nous pouvons utiliser $z^n - 1 = 0$, pour lequel la meilleure estimation est $|\hat{z}| \leq 1$. Puis, à l'aide de la substitution $x = 2z$, nous obtenons

$$\begin{aligned} x^n - 2^n = 0 \quad &\Longleftrightarrow \quad (2z)^n - 2^n = 0 \\ &\Longleftrightarrow \quad z^n - 1 = 0 \\ &\Longrightarrow \quad |z| \leq 1 \\ &\Longleftrightarrow \quad |x| \leq 2 \end{aligned}$$

1.11.2 Méthode de Newton-Maehly

Le problème traité ici est encore celui de trouver toutes les racines d'un polynôme $f(x)$. Supposons de connaître une valeur $\hat{x}$ telle que $f(\hat{x}) = 0$. Comme nous supposons que f n'a pas de racines multiples, pour chercher les autres nous posons

$$g(x) = \frac{f(x)}{x - \hat{x}}$$

Ainsi, les racines de g seront toutes les racines de f sauf $\hat{x}$. Puis, en utilisant la méthode de Newton pour calculer un zéro de g, nous allons trouver une racine ultérieure de f. Et ainsi de suite, jusqu'à la découverte de toutes les valeurs cherchées.
Mais à ce point un nouveau problème se présente: $\hat{x}$ n'est pas forcément rationnel. Du coup, les coefficients de g sont erronés! L'idée de Maehly, pour contourner ce problème, consiste à appliquer la méthode de Newton à g sans calculer ses coefficients.

Soient $\xi_1, \ldots, \xi_\ell$ des racines trouvées d'un certain polynôme f. Considérons

$$g(x) = \frac{f(x)}{(x - \xi_1) \cdots (x - \xi_\ell)}.$$

Par les propriétés du logarithme nous avons

$$\ln g(x) = \ln f(x) - \sum_{i=1}^{\ell} \ln(x - \xi_i).$$

Et, en dérivant cette dernière égalité, nous obtenons

$$\frac{g'(x)}{g(x)} = \frac{f'(x)}{f(x)} - \sum_{i=1}^{\ell} \frac{1}{x - \xi_i}.$$

D'où la formule

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k) - f(x_k) \sum\limits_{i=1}^{\ell} \frac{1}{x_k - \xi_i}}$$

qui reflète la méthode de Newton appliquée à g et qui livre une suite qui converge vers une solution de $f(x) = 0$ différente de ξ_i, $i = 1, \ldots, \ell$, s'il y en a une.

1.11.3 Suites de Sturm

Cette méthode, utilisée pour les racines réelles, est plus dans l'esprit de l'algorithme de Buchberger, par rapport à celles vues précédemment.

Définition 14. *Une suite finie de polynômes $(f_0, \ldots, f_n)$ à coefficients réels ou rationnels est une* ***suite de Sturm****, si pour $x^* \in \mathbb{R}$ elle satisfait*

1. $f_0'(x^*)f_1(x^*) < 0$, *si* $f_0(x^*) = 0$;
2. $f_{j-1}(x^*)f_{j+1}(x^*) < 0$, *si* $f_j(x^*) = 0$ *pour* $1 \leq j \leq n-1$;
3. $f_n(x)$ *ne change pas de signe.*

Théorème 18. *Soit $(f_0, \ldots, f_n)$ une suite de Sturm. Notons par $\omega(x)$ le nombre de changements de signes de $(f_0(x), \ldots, f_n(x))$. Alors la fonction f_0 possède exactement $\omega(b) - \omega(a)$ racines réelles dans l'intervalle $[a, b]$.*

Démonstration. Soit $\omega : \mathbb{R} \longrightarrow \mathbb{Z}$ définie comme dans le théorème. Supposons que nous connaissions un x^* tel que pour un certain i, $0 \leq i \leq n-1$, $f_i(x^*) = 0$ (f_n ne change pas de signe, par hypothèse). Analysons ce qui se passe dans un voisinage de x^*. Les tableaux suivants indiquent les signes des f_i dans un tel voisinage. Soit $\varepsilon > 0$ petit. Pour $1 \leq i \leq n-1$ nous pouvons avoir

	$x^* - \varepsilon$	x^*	$x^* + \varepsilon$
f_{i-1}	$+$	$+$	$+$
f_i	$\pm$	0	$\mp$
f_{i+1}	$-$	$-$	$-$

ou

	$x^* - \varepsilon$	x^*	$x^* + \varepsilon$
f_{i-1}	$-$	$-$	$-$
f_i	$\pm$	0	$\mp$
f_{i+1}	$+$	$+$	$+$

Ici, le nombre de changements de signe de la colonne de $x^* - \varepsilon$ est identique à celui de la colonne de $x^* + \varepsilon$. Pourtant la différence entre le nombre de changements de signe sur ces deux colonnes est en tout cas nulle. Par contre pour $i = 0$ les tableaux sont

	$x^* - \varepsilon$	x^*	$x^* + \varepsilon$
f_0	$+$	0	$-$
f_1	$+$	$+$	$+$

ou

	$x^* - \varepsilon$	x^*	$x^* + \varepsilon$
f_0	$-$	0	$+$
f_1	$-$	$-$	$-$

et dans les deux cas la susdite différence vaut 1. Ceci se produit pour de chaque racine de f_0. Il s'ensuit que $\omega(b) - \omega(a)$ compte effectivement le nombre de racines de f_0 comprises dans l'intervalle $[a, b]$. $\square$

La mise en pratique de ce théorème se fait comme suit.
Soit $f(x) \in \mathbb{R}[x]$ ou $\mathbb{Q}[x]$. Supposons que f n'a que des racines simples. Pour trouver une suite de Sturm, nous faisons:

1. $f_0(x) := f(x)$;
2. $f_1(x) := -f'(x)$;
3. $f_{k+1}(x)$ est, pour $k \geq 1$, le polynôme unitaire qui reste de la division euclidienne de f_{k-1} par f_k;[15]
4. nous nous arrêtons lorsque nous trouvons $f_n(x) \equiv 1$.

La suite $(f_0(x), \ldots, f_n(n))$ qui s'ensuit est bien une suite de Sturm, car si deux termes consécutifs sont nuls, alors toutes les valeurs qui les précèdent le sont aussi. Mais ceci est impossible, car f_0 et f_1 ne peuvent pas s'annuler en même temps, vu que f n'a pas de racines multiples.

Pour la recherche des racines réelles de f, une fois trouvés les intervalles dans lesquelles elles se situent, nous pouvons procéder par bissection.

1.11.4 Racines complexes

Pour trouver toutes les racines d'un polynôme, y compris les racines complexes, nous avons deux méthodes à disposition.
La première possibilité est d'utiliser la méthode de Newton ou de Newton-Maehly. La deuxième est de s'aider par des suites de Sturm. Regardons cette dernière.

Soit $z \in \mathbb{C}$. Ecrivons $z = x + iy$, où x et $y \in \mathbb{R}$. Soit $f(z) \in \mathbb{C}[z]$. Alors

$$f(z) = f(x + iy) = g(x, y) + ih(x, y)$$

avec g et $h \in \mathbb{R}[x, y]$. De plus, si $f \in \mathbb{R}[x]$, alors $g(x, y) = \hat{g}(x, y^2)$ et $h(x, y) = \hat{h}(x, y^2)y$[16].
Une fois développé f avec cette méthode, pour résoudre $f(z) = 0$, nous pouvons appliquer l'algorithme de Buchberger soit au système

$$\begin{cases} g(x, y) = 0 \\ h(x, y) = 0 \end{cases}$$

soit aux systèmes

$$\begin{cases} \hat{g}(x, y^2) = 0 \\ \hat{h}(x, y^2) = 0 \end{cases} \quad \text{ou} \quad \begin{cases} \hat{g}(x, 0) = 0 \\ y = 0 \end{cases}$$

selon les situations.

15. $f_{k-1}(x) = g_k(x)f_k(x) - c_k^2 f_{k+1}(x)$, avec $c_k \neq 0$ de manière que le coefficient dominant de f_{k+1} soit égal à 1.
16. $z = x + iy$, $z^2 = (x^2 - y^2) + i(2x)y$, $z^3 = (x^3 - 3xy^2) + i(3x^2 - y^2)y$, ...

Bibliographie

W.W. Adams & P. Loustaunau (1994), *An Introduction to Gröbner Bases*, American Mathematical Society, Graduate Studies in Mathematics, Vol.3 [MA 16/258]

B. Buchberger (1965), *Ein Algorithmus zum Auffinden der Basiselemente des Restklassenrings nach einem nulldimensionalen Polynomideal*, Ph. D. Thesis, Innsbruck

B. Buchberger (1979), *A criterion for detecting unnecessary reduction in the construction of Gröbner bases*, *Lecture Notes in Computer Science*, Vol. 72, pp. 3-21

B. Buchberger (1985), *Gröbner Bases: An Algorithmic Method in Polynomial Ideal Theory.* In Bose, N. (ed.), *Multidimension Systems Theory*, pp. 184-232, Reidel Publ. Comp., Dordrecht

D. Eisenbud (1965), *Commutative Algebra with a View Toward Algebraic Geometry*, Springer-Verlag, New York [MA 16/265]

R. Fröberg (1997), *An Introduction to Gröbner Bases*, John Wiley & Sons, Chichester [MA 16/319]

W. Gröbner (1949), *Moderne Algebraische Geometrie*, Springer-Verlag, Wien und Innsbruck [MA 14/55]

D. E. Knuth (1969), *The Art of Computer Programming*, Vol. 2 - Seminumerical Algorithms, Addison-Wesley, 2nd edition 1981 [MA 68/85]

M. Marden (1966), *Geometry of Polynomials*, American Mathematical Society, Providence, Rhode Island, 2nd edition [MA 30/54]

H. M. Möller & H. J. Stetter (1995), *Multivariate polinomial equation with multiple zeros solved by matrix eigenproblems*, Nummer. Math., Vol. 70, pp. 311-329

F. Pauer & M. Pfeifhofer (1988), *The theory of Gröbner bases*, L'Enseignement Mathématique, Vol. 34, pp. 215-232

H. J. Stetter (1996), *Computational Nonlinear Algebra*, Lectures (SoS 1996) et manuscrit non publié