Algorithmes modulaires de calcul des réductions de Hermite et de Smith

Salah LABHALLA Marrakech
Henri LOMBARDI Besançon
Roger MARLIN Nice

Résumé : Nous étudions dans cet article des algorithmes de calcul des réductions de Hermite et de Smith. Tout d'abord, nous étudions le problème du calcul des réductions de Hermite et de Smith pour les matrices non régulières, supposant connu le cas des matrices régulières. Nous discutons ensuite les méthodes modulaires.

Introduction

Nous considérons un anneau principal discret à division explicite **A** . Nous appelons **F** le corps de fractions de **A** .
A priori, les problèmes d'algèbre linéaire sur **F** sont plus faciles à traiter sur **F** que sur **A** .
Nous supposons que les déterminants sont «aisément» calculables dans **A** . Le fait de savoir calculer «aisément» (par exemple «en temps polynomial») les déterminants dans **A** permet de résoudre «aisément» les problèmes d'algèbre linéaire en dimension finie sur **F** , uniquement avec des calculs dans **A** .
Le fait que les déterminants dans **A** soient aisés à calculer est a priori une condition indispensable pour espérer calculer en un temps raisonnable les réduites de Hermite et Smith d'une matrice à coefficients dans **A** .
Les problèmes d'algèbre linéaire en dimension finie sur **A** se ramènent à des calculs de réductions de Hermite ou de Smith et à des calculs de produits de matrices à coefficients dans **A** . Par réduction d'une matrice, nous entendons : calcul de la réduite et calcul de la ou des matrices unimodulaires de changement de base. En particulier la solution des systèmes d'équations linéaires à coefficients et inconnues dans **A** est entièrement claire à partir de la réduction de Hermite des matrices.
Les méthodes «brutales», qui recopient les preuves explicites d'existence de la réduction de Hermite ou de celle de Smith, conduisent en pratique à une explosion de la taille des objets de l'anneau **A** manipulés par l'algorithme. En outre le nombre d'opérations arithmétiques dans l'anneau **A** est a priori mal controlé.

Dans cet article, nous éclairicissons quelques points concernant les algorithmes séquentiels qui évitent l'explosion de la taille des objets intermédiaires dans le calcul des réductions de Hermite et de Smith.

Dans une première partie, nous donnons quelques généralités concernant les réductions de Hermite et de Smith.

Dans la deuxième partie, nous étudions le problème du calcul des réductions de Hermite et de Smith pour les matrices non régulières, en supposant qu'a déjà été établie une méthode de calcul efficace pour le cas des matrices régulières. Dans toute la littérature sur le sujet, on ne met pas en évidence qu'on peut récupérer, à partir du cas non singulier, la réduction de Hermite et celle de Smith dans tous les cas (sans modification notable de la complexité du calcul). Certains auteurs font comme si c'était trivial (cf. [KKS]), d'autres signalent que ça semble un problème non résolu

(cf. notamment [HM], dernier en date de ceux que nous citons), d'autres enfin traitent quelques cas particuliers. En fait ce n'est pas trivial, mais ce n'est pas bien difficile.

Dans la troisième partie, nous étudions les méthodes modulaires, délà exposées avec différentes variantes par Frumkin, Domich-Kannan-Trotter, Schrijver, Iliopoulos et Haffner-MacCurley. Nous donnons un exposé simple de la méthode développée par Iliopoulos ([Ili1] et [Ili2]).

Grâce aux algorithmes développés dans la deuxième partie, il est maintenant clair que les méthodes modulaires permettent de traiter entièrement la réduction de Hermite et la réduction de Smith dans tous les cas (et non pas seulement la réduction de Hermite dans le cas non singulier, et la réduite de Smith dans le cas général). Ceci conduit à des algorithmes de complexité séquentielle polynomiale dans le cas des anneaux Z et $F[X]$ où F est un corps fini.

Pour les matrices à coefficients polynomiaux, nous renvoyons à [LLM].

1) Généralités sur les réduites de Hermite et de Smith

Définitions et notations

Nous nous situons dans le cadre général de la théorie constructive des anneaux principaux discrets (cf. [MRR] chap. IV, V) à division explicite. Un *anneau principal discret à division explicite* (au sens constructif) est un anneau intègre donné dans une présentation où :

l'anneau est discret : le test d'égalité est explicite,

l'anneau est explicite : les lois de compositions sont explicites,

l'anneau est explicitement de Bezout : on sait calculer à partir de a et b arbitraires, des éléments u, v, c, d, g avec :

$$u\,a + v\,b = g\,,\ g\,d = a,\ \ g\,c = b,$$

ce qui donne une matrice carrée de déterminant 1 avec :

$$(a,b)\begin{pmatrix} u & -c \\ v & d \end{pmatrix} = (g,0),$$

l'anneau est à division explicite : le test de divisibilité, et, en cas de réponse positive, le calcul du quotient, sont explicites, en particulier on a un test d'inversibilité et en cas de réponse positive, un calcul explicite de l'inverse,

toute suite décroissante pour la divisibilité contient deux termes consécutifs associés (c.-à-d. divisibles l'un par l'autre) : ceci est la condition de noetheriannité exprimée constructivement.

Tant que n'est pas envisagé le problème de la réduite de Smith, la condition de noetheriannité n'intervient aucunement. Si on supprime cette exigence, on dira que l'anneau $\mathbf{A}$ est un *anneau de Bezout explicite*.

Nous considérons donc un anneau de Bezout explicite $\mathbf{A}$ et une matrice M de type $s \times r$ (s lignes et r colonnes) à entrées dans $\mathbf{A}$. Nous appelons $\mathbf{F}$ le corps de fractions de $\mathbf{A}$. Nous notons :

L le module libre $\mathbf{A}^s$,

$e_1, \ldots, e_s$ la base canonique de L,

pour $k = 1, \ldots, s$; F_k est le sous-$\mathbf{A}$-module de L engendré par $e_k, \ldots, e_s$

$\Pi_k : L \longrightarrow \mathbf{A}$ la k-ème forme coordonnée

$V_1, \ldots, V_r$ les vecteurs colonnes de M (considérés comme des éléments de L),

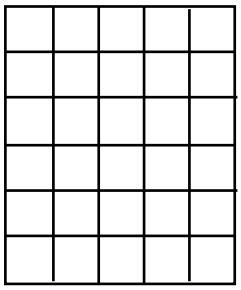
E le sous-$\mathbf{A}$-module de L engendré par $V_1, \ldots, V_r$

$E_k = E \cap F_k$

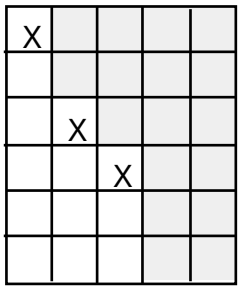
$E'_k = \Pi_k(E_k)$ qui est un idéal de **A**

Nous dirons qu'une matrice carrée est **unimodulaire** si son déterminant est inversible dans l'anneau **A** (il revient au même de dire que la matrice est inversible dans l'anneau des matrices carrées $\mathbf{M_r}(\mathbf{A})$).

Une **réduite de Hermite** de la matrice M est par définition une matrice M' de mêmes dimensions que M, dont les vecteurs colonnes engendrent le même sous-**A**-module E et qui de plus est «sous-triangulaire» au sens que les colonnes de M' successives contiennent de plus en plus de zéros au dessus de la première entrée non nulle. (voir figure).



matrice M



réduite de Hermite M'

Les croix représentent des entrées non nulles, les **pivots** de la matrice M', et la partie grisée représente des zéros : le nombre de zéros au dessus de la première entrée non nulle est strictement croissant tant que la colonne n'est pas entièrement nulle.

Une **réduction de Hermite** de la matrice M est donnée par une réduite de Hermite M' de M et une matrice unimodulaire P de type $r \times r$ vérifiant : M.P = M'

u	0	**-c**	0	0
0	1	0	0	0
v	0	**d**	0	0
0	0	0	1	0
0	0	0	0	1

Une réduction de Hermite de M peut être calculée en multipliant M à droite par des «matrices de Bezout» (une partie 2×2 du type Bezout, et le restant égal à la matrice identité : voir figure, avec $ud + vc = 1$) de manière à rendre nulles les entrées convenables les unes après les autres. Par exemple la matrice ci-contre, par multiplication à droite, produit une manipulation des colonnes 1 et 3, qui permet de remplacer, sur une ligne donnée, les coefficients a et b par g et 0. Dans l'exemple dessiné au dessus, on peut obtenir M' à partir de M en la multipliant par 9 matrices de Bezout successives.

Le problème qui nous préoccupe ici est que ce calcul est en général trop coûteux, parce que chacune des matrices de Bezout successives dépend des précédentes et qu'en conséquence sa taille est mal controlée.

Remarque : Une **manipulation élémentaire de colonnes** est un échange de colonnes ou le remplacement de la $i^{ème}$ colonne C_i par $C_i + a\,C_j$ avec $j \neq i$. Une **matrice de manipulation élémentaire** est obtenue en faisant subir une manipulation élémentaire de colonne à la matrice I. Faire une manipulation élémentaire de colonnes (resp. de lignes) sur une matrice revient à la multiplier à droite (resp. à gauche) par une matrice de manipulation élémentaire. Dans un anneau euclidien, les matrices de Bezout peuvent s'écrire comme produit de matrices de manipulations élémentaires. On en déduit que dans un anneau euclidien une matrice unimodulaire peut s'écrire comme produit d'une matrice unimodulaire diagonale et de matrices de manipulations élémentaires.

Une **réduction de Smith** de la matrice M est donnée par deux matrices unimodulaires L et C de types $s \times s$ et $r \times r$ vérifiant :

L.M.C = Δ est une matrice dont les seuls coefficients non nuls sont les éléments diagonaux $d_1, \ldots, d_m$ (où $m = \inf(r,s)$), avec d_i divise d_{i+1} pour $i = 1, \ldots, m-1$.

La matrice Δ est appelée une réduite de Smith de M. Elle est presque entièrement déterminée par M : le produit $d_1 \ldots d_i$ est un pgcd des mineurs d'ordre i de M. Donc chaque d_i est bien déterminé modulo la multiplication par un inversible.

Alors que le calcul d'une réduction de Hermite permet de discuter complètement la solution du système d'équations linéaires correspondant à la matrice considérée, le calcul d'une réduction de Smith donne des renseignements de nature géométrique plus fins concernant le **A**-module engendré par les vecteurs colonnes.
On peut calculer une réduction de Smith de M comme suit. On calcule une réduction de Hermite M_1 par manipulations élémentaires de colonnes, puis une réduction de Hermite de la transposée de M_1 par manipulations élémentaires de lignes de M_1, ce qui donne une matrice M_2. On recommence le processus jusqu'à obtenir une matrice diagonale. On termine ensuite facilement le passage de la forme diagaonale à la forme de Smith.

Remarque: dans l'anneau de Bezout non noethérien de tous les entiers algébriques, toute matrice admet une réduction de Smith. Il semble qu'on ignore toujours si une réduction de Smith existe pour toute matrice dans tout anneau de Bezout. (Kaplanski [Kap] montre que dans un anneau de Bezout toute matrice admet une réduction de Smith si et seulement si toute matrice triangulaire 2×2 en admet une)

Algèbre linéaire sur **A** et algèbre linéaire sur **F**

Nous supposons que les déterminants sont «aisément» calculables dans **A**. Par exemple nous supposons que la méthode de Bareiss ou celle de Berkovitz-Samuelson ne produit pas d'explosion du temps ce calcul, ce qui revient grosso modo à dire d'une part que les opérations d'addition, soustraction, multiplication sont «aisées», et d'autre part que les déterminants sont convenablement majorés en taille (pour la méthode de Bareiss, il faut aussi que les divisions exactes sont «aisées»). (cf. [Bar], [Ber], [Sam]).
En particulier rappelons une conséquence de l'algorithme de Berkovitz, dans le cas où les éléments de **A** sont codés de manière unique, ou au moins s'ils possèdent une représentation réduite calculable en temps polynomial : si les opérations arithmétiques dans **A** sont en temps polynomial et si les coefficients des polynomes caractéristiques sont polynomialement majorés dans **A** alors le calcul de ces coefficients (et donc en particulier le calcul des déterminants) est en temps polynomial dans **A**.
Le fait de savoir calculer «aisément» les déterminants dans **A** permet de résoudre «aisément» les problèmes d'algèbre linéaire en dimension finie sur **F**, uniquement avec des calculs dans **A**. Par contre les problèmes d'algèbre linéaire dans **A** sont a priori plus difficiles.

Dans la suite nous parlons de « triangulation dans **F** d'une matrice M » lorsque nous calculons des matrices M' et P avec :

$M' = M.P$, P inversible dans **F** *mais pas nécessairement dans* **A**, et M' sous forme triangulaire de Hermite comme expliqué au paragraphe précédent.

En pratique, cette triangulation dans **F** se fait par la méthode de Bareiss ou celle de Berkovitz-Samuelson, méthodes qui impliquent uniquement des calculs dans l'anneau des entrées de M. Toutes les entrées non nulles de M' sont alors égales à des déterminants extraits de M.
En résumé : une «triangulation dans **F** » ne signifie pas que les calculs ont lieu dans **F** (ils ont lieu dans **A**) mais que la matrice inversible P est a priori seulement inversible dans **F**.

Une caractérisation algébrique «abstraite» des réduites de Hermite

Les résultats suivants sont faciles, et sous une forme voisine, toujours présents dans la littérature sur le sujet.

Proposition 1 : On reprend les notations du début du a).
On considère un entier $t \leq r$ et une matrice M' dont les t premières colonnes sont non nulles et les $r-t$ dernières nulles. On note $W_1, \ldots, W_t$ les t premiers vecteurs colonnes de M', k_i (i=1, ... , t) le numéro de la première entrée non nulle de W_i, w_i l'entrée en question.
i) Si M' est une réduite de Hermite de M, les conditions suivantes sont vérifiées :
- la suite k_i (i=1, ... , t) est strictement croissante
- w_i (i=1, ... , t) engendre l'idéal E'_{k_i}
- $E'_k = \{0\}$ pour tout k distinct des k_i

ii) Réciproquement, si les W_i sont dans E et si les trois conditions ci-dessus sont vérifiées, la matrice M' est une réduite de Hermite de M.

Corollaire : Etant données deux réduites de Hermite M' et M" d'une même matrice M de type $s \times r$, il existe une matrice Q de type $r \times r$, sous-triangulaire et avec des inversibles sur la diagonale telle que M'.Q = M". En particulier toute réduite de Hermite de M provient d'une réduction de Hermite de M (c.-à-d. est de la forme M.P avec P unimodulaire).

Remarque sur les coefficients pivots de la réduite de Hermite. Le produit à droite par une matrice unimodulaire ne change pas le pgcd des mineurs d'ordre m extraits sur m lignes fixées d'une matrice. En conséquence, nous avons :

w_1 = pgcd des coefficients de la ligne numéro k_1 de M (à un inversible près)
$w_1 w_2$ = pgcd des mineurs d'ordre 2 extraits sur les lignes numéro k_1 et k_2 de M
$w_1 w_2 w_3$ = pgcd des mineurs d'ordre 3 .... etc ...

Remarque sur la triangulation dans F de la matrice M.
Une triangulation de M dans **F** par manipulations de colonnes et sans échanges de lignes n'est rien d'autre que le calcul d'une réduite de Hermite de la matrice M, lorsque l'anneau considéré est le corps **F** (qui, étant un corps explicite, est a fortiori un anneau de Bezout explicite). Notez que la triangulation dans **F** de la matrice M fournit la «forme» de la réduite de Hermite dans **A**, c.-à-d. le rang de la matrice et les numéros k_i des lignes des pivots.

Cas des entiers naturels

Lorsque l'anneau **A** est Z, on peut définir, parmi toutes les réduites de Hermite d'une matrice M, une **réduite normale**. C'est celle où les pivots sont positifs, et les coefficients à gauche du pivot w_i sont réduits modulo w_i (on peut choisir quelle réduction modulo w_i on utilise, ce qui modifie la réduite normale). Lorsque le type de «réduction modulo» est fixé, la réduite normale est unique, et elle est alors de taille «raisonnable».
Ce n'est pas avant 1976 (cf. les références [Fru x]) qu'on a su résoudre en temps polynomial les systèmes d'équations linéaires en entiers. Les travaux de Frumkin utilisent des méthodes modulaires et donnent de fait la possibilité de trianguler à la Hermite ou de diagonaliser à la Smith des matrices à coefficients entiers, même si l'auteur n'a pas bien su “vendre sa marchandise”, ce qui fait qu'on ne lui attribue pas la paternité à laquelle il a droit.
En 1979, Kannan et Bachem [KB] donnent un algorithme qui calcule la réduite normale en évitant l'explosion exponentielle de la taille des coefficients intermédiaires, sans utiliser les méthodes modulaires.

Algorithmes efficaces: méthodes et résultats connus

Nous examinons très rapidement la littérature concernant la réduction en temps polynomial des matrices à coefficients entiers. Tous les algorithmes tournent également en temps polynomial lorsque l'anneau des coefficients est un anneau $F[X]$ où F est un corps fini.

Méthodes directes (Kannan-Bachem [KB] et Chou-Collins [CC])
Soit M une matrice carrée régulière de rang r à coefficients entiers. Notons $M(k)$ la matrice mineure principale de M d'ordre k, et $m_{i,j}$ le coefficient de M en position (i,j). Il faut éventuellement, au moyen d'une triangulation péalable dans Q, permuter préalablement les colonnes de manière que toutes les matrices mineures principales soient non singulières
L'algorithme proposé dans [KB] transforme successivement $M(k)$ (pour $k = 2,...,r$) en sa forme normale de Hermite en deux étapes. .

Pour tous les k de 1 à $r–1$ faire

> Etape 1 : Annulation des $m_{i\ k+1}$ (successivement pour $i = 1,\ldots,k$) en multipliant la matrice M par des matrices de Bezout opérant sur les colonnes i et $k+1$.
> Après cette étape la matrice $M(k+1)$ est sous forme de Hermite non nécessairement normale.
>
> Etape 2 : Normalisation des coefficients à gauche des pivots de manière que sur chaque ligne où se trouve un pivot, à sa gauche, les éléments soient positifs et plus petits que le pivot.

Pendant tout l'algorithme la taille des coefficients reste majorée par n^4 Log($n.||M||$) où $||M|| = \max_{i,j}\{ \log_2 (|m_{i\,j}|+1) \}$.

Sans modification majeure de cet algorithme, Chou et Collins ont montré que la taille des coefficients reste majorée par n^2 Log($n.||M||$) : la seule modification porte sur l'ordre dans lequel on normalise les coefficients à gauche des pivots, au lieu que ce soit par ligne de la gauche vers la droite et du haut vers le bas, c'est par colonne de la droite vers la gauche et du bas vers le haut.

Méthode modulaire
L'origine de cette méthode est dans les articles de Frumkin ([Fru1], [Fru2], [Fru3]). Le tribut ne lui est pas en général justement rendu. Cependant, la méthode n'a été exposée en tant que méthode de manipulation de matrices que plus tard, dans les articles [Ili1], [Ili2], [DKT] et dans le livre [Sch]. Il semble bon de signaler qu'Iliopoulos a exposé l'idée de base de sa méthode dans un colloque antérieur à toutes ces publications. L'idée essentielle est de faire tous les calculs modulo un déterminant extrait de rang maximum. Les principes de base sont bien expliqués dans [Sch]. Il semble que les articles d'Iliopoulos contiennent quelques améliorations intéressantes, que nous reprenons dans la partie 3). L'exposé dans les articles [Ili] n'est cependant pas limpide. Nous recommanderions plutôt l'excellent livre de Schrijver. L'article [DKT] contient une variante interessante. Enfin l'article [HM] propose encore une amélioration par rapport à [DKT].

Le cas non régulier. Dans toute la littérature sur le sujet, on ne met pas en évidence qu'on peut récupérer, à partir du cas non singulier, la réduction de Hermite et celle de Smith dans tous les cas (sans dommage pour la complexité du temps de calcul). Certains auteurs font comme si c'était trivial (cf. [KKS]), d'autres signalent que ça semble un problème non résolu (cf. notamment [HM], dernier en date de ceux que nous citions), d'autres enfin traitent quelques cas particuliers. En fait ce n'est pas trivial, mais ce n'est pas bien difficile. C'est l'objet de notre partie (2).

2) La réduction du cas singulier au cas non singulier

De la réduite de Hermite générale à la réduction de Hermite

Certains algorithmes de calculs de réduites de Hermite fonctionnent sans restriction aucune sur la matrice. Voyons comment le fait de savoir calculer rapidement une réduite de Hermite (pour une matrice M arbitraire) permet de calculer rapidement une réduction de Hermite.
Considérons en effet la matrice N obtenue à partir de M en «collant» en dessous une matrice identité de type $r \times r$ et calculons une réduite de Hermite N' de N. Elle est constituée d'une matrice M' de type $s \times r$ avec collée en dessous une matrice P de type $r \times r$. Comme les colonnes de N et celles de N' engendrent le même sous-**A**-module de $\mathbf{A}^{s+r}$ il existe deux matrices U et V de type $r \times r$ à entrées dans **A** telles que : $N\,U = N'$ et $N'\,V = N$. D'où on tire facilement que $U = P$ et $P\,V = I_r$. Donc $M\,P = M'$, avec P unimodulaire. Et la matrice M' est bien de la forme voulue.

$$\begin{bmatrix} M \\ I_r \end{bmatrix} \bullet \boxed{U} = \boxed{N'} = \begin{bmatrix} M' \\ U \end{bmatrix} \quad , \quad \boxed{N'} \bullet \boxed{V} = \begin{bmatrix} M'.V \\ U.V \end{bmatrix} = \begin{bmatrix} M \\ I_r \end{bmatrix}$$

De la réduite de Hermite non singulière à la réduction de Hermite générale

Certains algorithmes de calcul des réduites de Hermite ne fonctionnent bien que dans le cas d'une matrice carrée non singulière. Dans ce paragraphe, nous indiquons comment, à partir de ce cas, on peut traiter le cas général

Premier cas, matrice carrée non singulière

Soit M' la réduite de Hermite de M, on cherche P, unimodulaire, tel que $M' = M\,P$.
Comme la solution est unique et automatiquement unimodulaire, elle se calcule par les algorithmes d'algèbre linéaire dans le corps des fractions **F**.

Deuxième cas, matrice de rang égal à son nombre de colonnes

Une triangulation dans **F** de la matrice M donne les numéros des lignes des pivots. On extrait de M la matrice N correspondant à ces lignes. On calcule la réduction de Hermite $N' = N\,P$, on a la réduction de Hermite de M en faisant $M' = M\,P$.

Troisième cas, matrice de rang égal à son nombre de lignes (cf.[KB])

La matrice M est une matrice de type $s \times r$, de rang s avec $r > s$.
Par triangulation dans **F** de la matrice, on détermine une permutation de colonnes qui permet de ramener s colonnes indépendantes en première position.

Nous supposons désormais que la matrice est de ce type.
(avec $\det(M_1) \neq 0$)

$$\boxed{M_1 \mid M_2}$$

On considère alors la matrice M_3 non singulière égale à :

M_1	M_2
0	I_{r-s}

Une réduite de Hermite de M_3 est une matrice
$T = M_3 . P$ avec P unimodulaire

T_1	0
U	T_2

Alors $T' =$ | T_1 | 0 | est une réduite de Hermite de M puisque $T' = M$.

Cas général

Supposons maintenant le rang de M strictement inférieur à r .
On commence, par une triangulation dans **F** de la matrice M (sans échange de lignes) à calculer les indices k_1 , …, k_t des lignes des pivots.
On décompose la matrice en blocs successifs de manière à séparer les lignes portant des pivots, des autres. Par exemple, avec $s = 8$, $r = 6$, $k_1 = 1$, $k_2 = 2$, $k_3 = 4$, $k_4 = 5$, $k_5 = 8$, on a les 5 blocs ci-contre.

$M =$

M_1
M_2
M_3
M_4
M_5

On conserve uniquement les blocs portant les pivots, et on obtient une matrice N qui rentre dans le troisième cas examiné :

$N =$

M_1
M_3
M_5

Soit alors $N.P = T'$ une réduction de Hermite de N. On décompose P en blocs verticaux de manière à faire apparaître dans le produit les blocs triangulaires $M_i.P_i$ sur la diagonale :

P_1	P_3	P_5	P_6

M_1
M_3
M_5

M_1P_1	0	0	0
M_3P_1	M_3P_3	0	0
			0

Montrons alors que $M.P = T''$ est une réduction de Hermite de M

Puisque P est unimodulaire, il suffit de montrer que la matrice T'' est de la forme voulue, ce qui revient à dire que des blocs tels que $M_2.P_3$ sont nuls.
Si $M_2.P_3$ n'était pas nulle, comme M_1P_3 = une matrice nulle, il y aurait une combinaison linéaire qui annulerait les colonnes de M_1 sans annuler celles de M_2. Cela revient à dire qu'il y a un hyperplan qui contient les lignes de M_1 mais ne contient pas toutes les lignes de M_2, c.-à-d. encore que la matrice obtenue en mettant M_1 au dessus de M_2 aurait un rang supérieur au rang de M_1, mais ceci contredit la triangulation de M dans **F**.

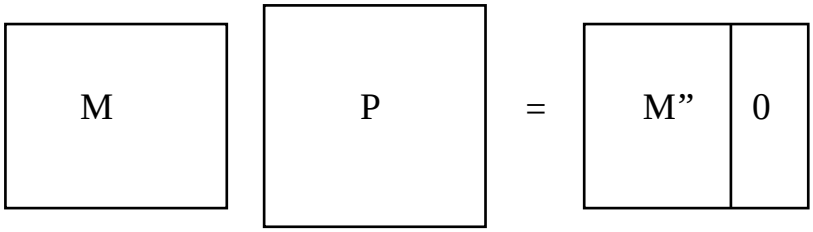
P_1	P_3	P_5	P_6

M_1
M_2
M_3
M_4
M_5

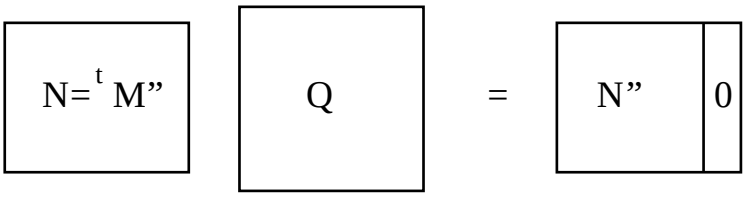
	0	0	0
	M_2P_3		
		0	0
			0

Des réductions de Hermite et de Smith non singulière à la réduction de Smith générale

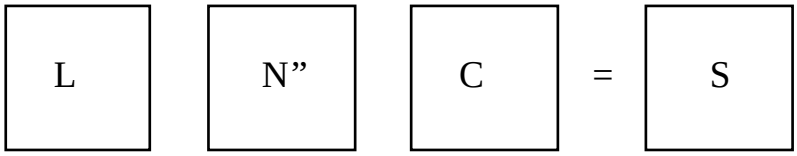
Nous supposons avoir un algorithme convenable (en temps polynomial par exemple) pour la réduction de Hermite et un autre pour la réduction de Smith dans le cas des matrices carrées non singulières. Nous obtenons la réduction de Smith en enchainant deux réductions de Hermite et une réduction de Smith non singulière. (nous ne rappelons ici pas la définition bien connue de la réduction de Smith)
Soit en effet M une matrice rectangle ou carrée singulière, et $M.P = M'$ une réduction de Hermite. Si M'' est la matrice extraite de M' en ne conservant que les colonnes non nulles, il est clair qu'il suffit de calculer une réduction de Smith de M''.



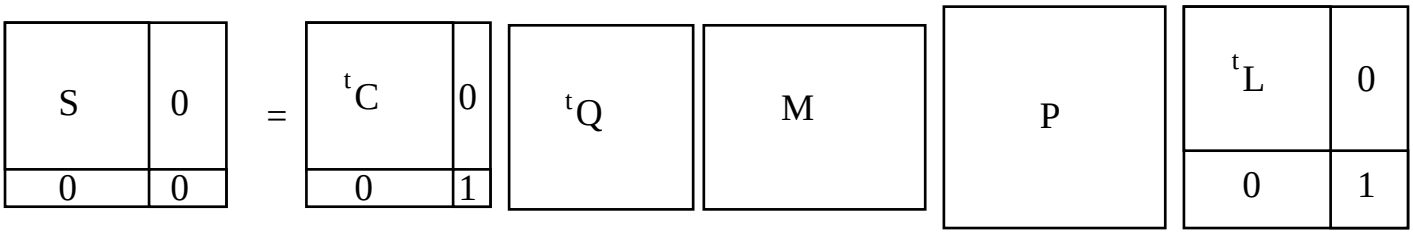
Si M” est carrée donc non singulière on sait par hypothèse calculer cette réduction de Smith (cela signifiait pour M que son rang était celui de son nombre de lignes).
Sinon, soit N la matrice transposée de M” et soit N.Q = N’ une réduction de Hermite de N . Soit enfin N” la matrice extraite de N’ en ne gardant que les colonnes non nulles.



Cette fois-ci, N” est non singulière, et on sait calculer une réduction de Smith.



Ce qui donne la réduction de Smith de M :

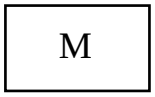
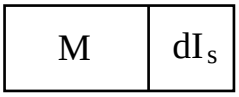


Notons enfin que si le rang de M est égal à son nombre de colonnes, on fait le calcul ci dessus avec la transposée de M ce qui évite la deuxième réduction de Hermite.

3) Les méthodes modulaires

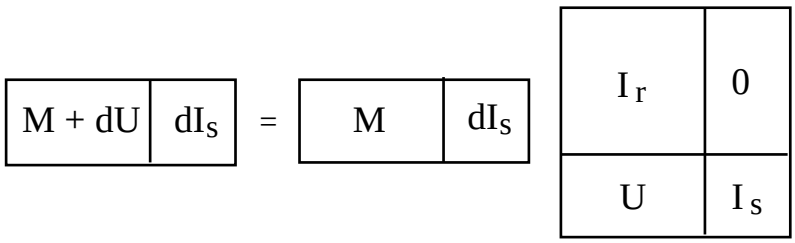
Remarques fondamentales (cf. [Sch] et [Ili1])

On suppose qu'on est dans un anneau euclidien (tel que Z ou F[X] pour un corps fini F) dans lequel on a un bon algorithme de division, et où les restes modulo un élément d ont une taille bien majorée en fonction de la taille de d . On désire calculer des réduites de Smith ou de Hermite de manière modulaire. Nous utilisons un langage «opératoire», et appelons **manipulation de colonnes** (resp. de lignes) **légitime** sur une matrice M , toute opération sur les colonnes (resp. les lignes) de M qui équivaut à multiplier la matrice à droite (resp. à gauche) par une matrice unimodulaire. En particulier les manipulations élémentaires sont légitimes.

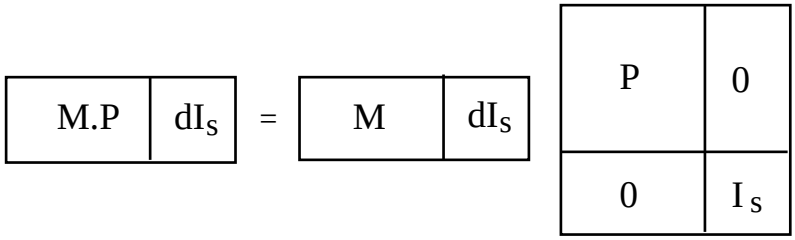
A toute matrice  de type $s \times r$ on associe la matrice $N =$ 

Alors :

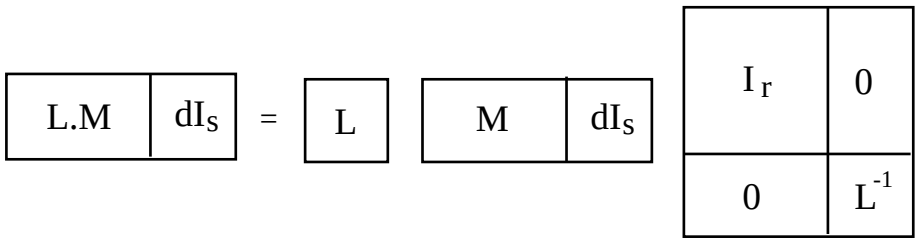
Remarque 1 : réduire des coefficients de M modulo d revient à opérer une manipulation de colonnes légitime sur N



Remarque 2 : opérer une manipulation de colonnes légitime sur M revient à opérer une manipulation de colonnes légitime sur N



Remarque 3 : Opérer une manipulation de lignes légitime sur M revient à opérer une manipulation de colonnes et une manipulation de lignes légitimes sur N .



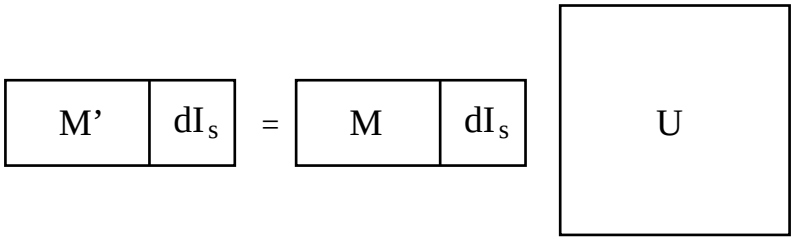
Remarque 4 : Si la colonne numéro j de M est nulle sauf sur la ligne numéro i , et si on remplace le coefficient non nul en question a par pgcd(a,d) , cette transformation peut être obtenue par manipulations légitimes de colonnes sur N.
En effet, en multipliant à droite N par une matrice de Bezout, on remplace sur la ligne numéro i le coefficient a en colonne numéro j , et le coefficient d en colonne numéro i+r (dans la partie $d.I_s$) respectivement par pgcd(a,d) et 0 . Il suffit ensuite de rétablir le d dans la partie $d.I_s$ en ajoutant d/pgcd(d,a) fois la colonne numéro j à la colonne numéro i+r .
Ceci peut être écrit sous forme d'un algorithme, en notant C_k la colonne numéro k de N .

```
Calculer  u  et  v  tels que  u a + v d = δ = pgcd(d,a)
```

$$(C_j, C_{i+r}) \leftarrow (C_j, C_{i+r}) \begin{pmatrix} u & -d/\delta \\ v & a/\delta \end{pmatrix}$$

$$(C_j, C_{i+r}) \leftarrow (C_j, C_{i+r}) \begin{pmatrix} 1 & d/\delta \\ 0 & 1 \end{pmatrix}$$

Définition : Nous dirons qu'une matrice M' est obtenue à partir d'une matrice M par manipulation d-modulaire légitime de colonnes, s'il existe une matrice unimodulaire U telle que :



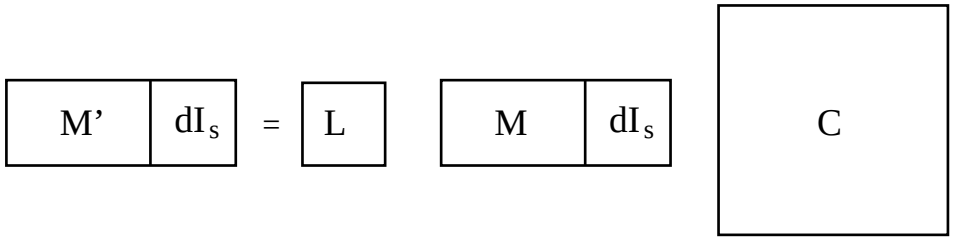
Nous appellerons méthode modulaire toute méthode qui combine des manipulations d-modulaires légitimes de colonnes et des manipulations légitimes de lignes.

D'après les remarques 1, 2, 3 et 4, on a :

Lemme 1 :

a) Les manipulations légitimes de colonnes, les réductions modulo d et le remplacement, dans une colonne ayant un seul coefficient non nul a , de ce coefficient a par pgcd(a,d), sont des manipulations d-modulaires légitimes de colonnes.

b) Si on transforme M en M' au moyen d'une méthode modulaire (modulo d), il existe des matrices unimodulaires L et C telles que



Réduite de Hermite, le cas d'une matrice carrée non singulière

Soit M carrée non singulière de type $r \times r$ de déterminant d et soit M_1 la comatrice de M (c.-à-d. la matrice à coefficients dans **A** telle que $M.M_1 = d.I_r$).
Soit N la matrice obtenue en collant $d.I_r$ à droite de M
Soit $T = M.P$ une réduction de Hermite de M (P est unimodulaire).

Alors la matrice $\left[\begin{array}{c|c} T & 0 \end{array}\right]$ est une réduite de Hermite de $\left[\begin{array}{c|c} M & dI_r \end{array}\right]$.

en effet : $\left[\begin{array}{c|c} T & 0 \end{array}\right] = \left[\begin{array}{c|c} M & dI_r \end{array}\right] \left[\begin{array}{c|c} P & -M_1 \\ \hline 0 & I_r \end{array}\right]$ (a)

Si maintenant la matrice $\left[\begin{array}{c|c} T' & 0 \end{array}\right]$ est une autre réduite de Hermite de $\left[\begin{array}{c|c} M & dI_r \end{array}\right]$ on obtient une égalité $\left[\begin{array}{c|c} T & 0 \end{array}\right] = \left[\begin{array}{c|c} T' & 0 \end{array}\right] \left[\begin{array}{c|c} P' & R \\ \hline Q & S \end{array}\right]$

On en déduit que $T'P' = T$ et donc que P’ est unimodulaire, puique T et T’ ont, à un inversible près, le même déterminant. En conclusion T’ est une réduite de Hermite de M .

Lemme 2 : Soit M carrée non singulière de type $r \times r$ de déterminant d et soit N la matrice obtenue en collant $d.I_r$ à droite de M . Toute réduite de Hermite de N fournit sur ses r premières colonnes une réduite de Hermite de M (et ses r dernières colonnes sont nulles).

On peut donc réduire la matrice N à une forme de Hermite en procédant ligne après ligne, en conservant la colonne n°r+j intacte tant que les lignes traitées sont de numéros 1, ..., $j - 1$. Cela permet de faire la plupart des calculs modulo d puisqu'une réduction modulo d sur les lignes non encore traitées revient à faire une manipulation élémentaire de colonnes sur la matrice N . Par contre, il semble impossible d'obtenir à coup sûr une réduite de Hermite de M en travaillant modulairement sur la seule matrice M , même si on prend la précaution d'aboutir systématiquement à des diviseurs de d sur la diagonale :

Cf. l'exemple de la matrice $\begin{pmatrix} 6 & 12 \\ 1 & 4 \end{pmatrix}$ de déterminant égal à 12 et pour laquelle une réduction modulo 12 donne une fausse réduite de Hermite, avec des diviseurs du déterminant sur la diagonale.
Signalons cependant que [DKT] donne une méthode qui permet d'éviter de coller une matrice identité à droite de la matrice M.

Algorithme modulaire de calcul d'une réduite de Hermite d'une matrice carrée non singulière.

1) Calculez le déterminant d de la matrice M. Appelez N la matrice obtenue en collant $d.I_r$ à droite de M.
2) Appliquez votre algorithme préféré de calcul d'une réduite de Hermite à la matrice N, en procédant ligne après ligne et en ayant soin de profiter au maximum des réductions modulo d autorisées par la présence des colonnes non encore modifiées de la matrice N. Gardez les r premières colonnes de la matrice obtenue.

En reprenant l'idée d'Iliopoulos [Ili1] pour minimiser le nombre de produits par des matrices de Bezout, on obtient en particulier, dans le cas des anneaux Z ou $F[X]$, l'algorithme suivant :

Algorithme à la Iliopoulos (réduite de Hermite à coefficients entiers, cas non singulier):

```
Calculer le déterminant d de la matrice M.
Réduire la matrice M modulo d . Vous obtenez une matrice M'.
Appelez N la matrice obtenue en collant d.I_r à droite de M'.
Pour i de 1 à r
{invariant de boucle : N est sous-triangulaire sur les lignes correspondant aux indices pour lesquels la boucle a été exécutée. Elle est obtenue à partir de la matrice initiale par manipulations élémentaires de colonnes, et ses colonnes de r+i+1 à 2r sont inchangées }
   | Si N(i,i) ≠ 0
   |    | m_i ← N(i,i)
   |    | m_{i+1} ← N(i,i+1)    (*)
   |    | calculer u et v tels que u.m_i + v.m_{i+1} = δ = pgcd(m_i,m_{i+1}))
   |    | (C_{i+1},C_i) ← (C_{i+1},C_i) ( v   -m_i/δ ) modulo d
   |    |                               ( u  m_{i+1}/δ )
   | finsi
   | { N(i,i) = 0 et l'invariant de boucle est vrai }
   | Calculer u(i+1…r+i) tels que
   |          Σ_{j=i+1…r+i} u(j) N(i,j) = pgcd(N(i,i+1…r+i)) = d_i
   | C_i ← C_i + Σ_{j=i+1…r+i} u(j).C_j modulo d   (**)
   | { N(i,i) est maintenant égal à d_i qui divise d }
   | Pour j de i+1 à r+i
   | { les N(i,k) pour les indices k exécutés de cette boucle sont nuls,
   | et l'invariant de boucle est vrai }
   |    | c ← N(i,j)/d_i
   |    | C_j ← C_j - c.C_i   modulo d    (***)
   | finpour
finpour
(*) sur cette ligne et les deux suivantes l'indice de colonne i+1 peut
être remplacé par n'importe quel j compris entre i+1 et r+i-1. En
particulier si N(i,j) = 0 pour un tel j , le calcul du pgcd et la
multiplication par la matrice de Bezout est remplacé par un simple
échange de colonnes
```

```
(**) sur la ligne i on met directement d_i sans effectuer le calcul
(***) sur la ligne i on met directement 0 sans effectuer le calcul
```

Remarque : Au cours de l'algorithme, les valeurs de $N(i,j)$ sont toutes calculées modulo d, ce qui permet de maitriser parfaitement leur taille. Cependant les éléments égaux à d sur la diagonale de $d.I_r$ sont réduits à 0 par une manipulation de colonnes et non par une réduction modulo d. En outre d'autres calculs de l'algorithme n'ont pas systématiquement lieu modulo d : cf. le calcul de u et v tels que $u.m_i + v.m_{i+1} = \delta = pgcd(m_i,m_{i+1})$)

Réduite de Smith, le cas général

On considère une matrice $s \times r$ arbitraire M à coefficients dans $\mathbf{A}$.
Une matrice rectangulaire est dite diagonale si ses seules entrées non nulles sont sur la diagonale.
Nous appellerons **forme diagonale de Smith** de la matrice M toute matrice diagonale qui peut être obtenue à partir de M par manipulations légitimes de lignes et de colonnes.
On peut passer d'une forme diagonale de Smith à la réduite de Smith par l'algorithme classique suivant:

```
Pour i de 1 à inf(r,s) – 1
   | Pour j de i+1 à inf(r,s)
   |    | a ← D(i,i)
   |    | b ← D(j,j)
   |    | Si b ne divise pas a
   |    |    | D(i,i) ← pgcd(a,b)
   |    |    | D(j,j) ← ppcm(a,b)
   |    | finsi
   | finpour
finpour
```

Cet algorithme est un «raccourci» pour un algorithme qui n'utilise que des manipulations légitimes de lignes et de colonnes. En effet, si $u\,a + v\,b = pgcd(a,b) = g$ on a :

$$\begin{pmatrix} a & b \\ 0 & b \end{pmatrix}\begin{pmatrix} u & -b/g \\ v & a/g \end{pmatrix} = \begin{pmatrix} g & 0 \\ bv & ba/g \end{pmatrix}$$

et donc :

$$\begin{pmatrix} 1 & 0 \\ -bv/g & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}\begin{pmatrix} u & -b/g \\ v & a/g \end{pmatrix} = \begin{pmatrix} g & 0 \\ 0 & ba/g \end{pmatrix}$$

ou si on préfère, en transposant :

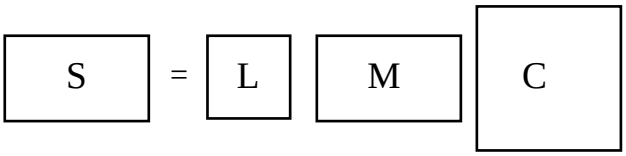
$$\begin{pmatrix} u & v \\ -b/g & a/g \end{pmatrix}\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}\begin{pmatrix} 1 & -bv/g \\ 1 & 1-bv/g \end{pmatrix} = \begin{pmatrix} g & 0 \\ 0 & ba/g \end{pmatrix}$$

La justification du calcul modulaire de la réduite de Smith repose sur le lemme suivant.

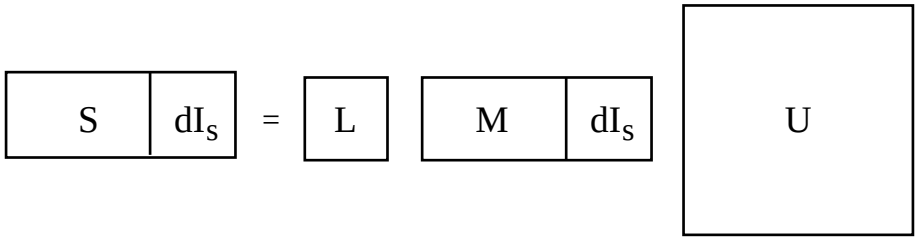
Lemme 3 : On considère une matrice $s \times r$ arbitraire M à coefficients dans $\mathbf{A}$. Soit d un pgcd de déterminants de mineurs de rang maximum extraits de M. Soit D une matrice diagonale obtenue à partir de M par manipulations d-modulaires légitimes de colonnes et par des manipulations légitimes de lignes. Si les éléments diagonaux de D divisent d, alors D contient au moins $inf(r,s) - t$ entrées égales à d sur la diagonale. On obtient alors une forme diagonale de Smith de M en remplaçant $inf(r,s) - t$ de ces entrées par zéro.

preuve> Soit S la réduite de Smith de la matrice M, notons $(d_1, \ldots, d_t)$ son type (la liste de ses éléments diagonaux non nuls, en ordre décroissant pour la divisibilité). Le produit des d_i divise d, en particulier chaque d_i divise d.

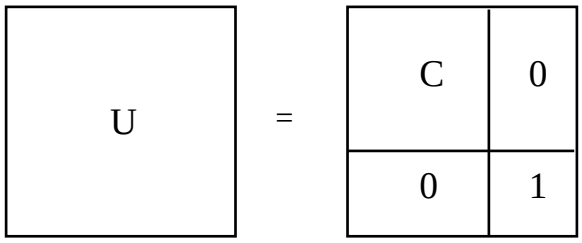
On a une égalité :



avec L et C unimodulaires, et donc aussi



avec



Soit S' la matrice déduite de S en remplaçant chaque 0 sur la diagonale par d .

Il est facile de voir que la réduite de Smith de $\begin{array}{|c|c|}\hline S & dI_s \\ \hline \end{array}$ est $\begin{array}{|c|c|}\hline S' & 0 \\ \hline \end{array}$ (on passe de la première à la seconde par manipulations légitimes de colonnes).

Par ailleurs, vues les remarques 1 à 4, la matrice $\begin{array}{|c|c|}\hline D & dI_s \\ \hline \end{array}$ est obtenue à partir de $\begin{array}{|c|c|}\hline M & dI_s \\ \hline \end{array}$ par manipulations légitimes de lignes et de colonnes.

Puisque les coefficients de D divisent d , on peut passer de $\begin{array}{|c|c|}\hline D & dI_s \\ \hline \end{array}$ à $\begin{array}{|c|c|}\hline D & 0 \\ \hline \end{array}$ par manipulations légitimes de colonnes.

En conclusion, on passe de $\begin{array}{|c|c|}\hline S' & 0 \\ \hline \end{array}$ à $\begin{array}{|c|c|}\hline D & 0 \\ \hline \end{array}$ par manipulations légitimes de lignes et de colonnes, et donc S' est la réduite de Smith de D .

Vu l'algorithme « forme diagonale de Smith ¾→ forme réduite de Smith » décrit ci-dessus, on ne diminue pas le nombre de coefficients égaux à d lorsqu'on passe de D à S'. Ceci termine la preuve du lemme. o

Algorithme modulaire de calcul d'une réduite de Smith d'une matrice arbitraire.

1) Déterminez le rang t de la matrice M et calculez un pgcd d de quelques mineurs extraits d'ordre t (un non nul suffit en théorie, mais deux non nuls en pratique peuvent permettre d'abaisser notablement d et donc le coût total du calcul).
2) Appliquez modulo d votre algorithme préféré de calcul d'une forme diagonale de Smith à la matrice M . Vous obtenez une matrice D .
3) Appliquez à D l'algorithme suivant :

```
Pour i de 1 à inf(r,s)
   | T(i,i) ← pgcd(d,T(i,i))
finpour
remplacer inf(r,s)—t fois la valeur d par zéro.
```

4) Passez de la forme diagonale à la réduite de Smith par votre algorithme préféré.

preuve de l'algorithme> claire d'après les lemmes 1 et 3 o

Réduction de Smith, le cas d'une matrice carrée non singulière

Soit M, matrice carrée $r \times r$ non singulière de déterminant d, et soit M_1 sa comatrice.
Supposons qu'on ait obtenu par une méthode modulaire (modulo d) la réduite de Smith D de M. On en déduit facilement qu'on a une égalité :

$$L.M.P = D + dU \quad \text{avec } L \text{ unimodulaire, } (L, P \text{ et } U \text{ carrées } r \times r)$$

Puisque D est la réduite de Smith de M, on a :

$$\det(M) = \det(D) \text{ (à un inversible près)}$$

On a aussi : $L.M.(P - M_1.L^{-1}.U) = D$
Et la matrice $P' = P - M_1.L^{-1}.U$ à coefficients dans **A** est donc unimodulaire.
Autrement dit encore D est la réduite de Hermite de $L.M$.

D'après les remarques 1 à 4 du début il est facile d'expliciter L au cours du calcul : essentiellement, on fait subir à la matrice I_r les mêmes manipulations de lignes que celles qu'on fait subir à M. Notez cependant que si on se ramène tout d'abord à une forme diagonale, le passage de la forme diagonale à la forme réduite cache inévitablement des manipulations de lignes, qu'il faut expliciter.
Il serait plus fatigant d'expliciter P et U au cours du calcul. Nous préfèrerons la méthode suivante.

Expliciter L^{-1} au cours du calcul (ou le calculer à partir de L à la fin).
Calculer P' par $P' = (M_1 L^{-1} D) / d$ (c'est un calcul d'algèbre linéaire dans **F**).

Réductions de Hermite et de Smith : le cas général

D'après les remarques de la partie 2), et puisque nous disposons d'algorithmes modulaires efficaces pour la réduite de Hermite et la réduction de Smith non singulière, nous disposons d'algorithmes efficaces pour les réductions de Hermite et de Smith dans le cas général.

Roger MARLIN
Département de Mathématiques
UFR des Sciences
Université de Nice
06034 NICE Cédex
FRANCE

Henri LOMBARDI
Laboratoire de Mathématiques
URA CNRS 741
UFR des Sciences et Techniques
Université de Franche-Comté
25030 BESANCON Cédex
FRANCE

Salah LABHALLA
Dép. de Mathématiques
Université de Marrakech
Bd de SAFI. BP S 15
MARRAKECH
MAROC

Bibliographie

[**Bar**] Bareiss E. H. : *Sylvester's Identity and Multistep Integer-Preserving Gaussian Elimination.* Math. Comp. 22 565-578 (1968) .

[**Ber**] Berkovitz S. J. : *On computing the determinant in small parallel time using a small number of processors.* Information Processing Letters 18 numéro 3 147-150 (1984) .

[**CC**] Chou T.-W., Collins G. : *Algorithms for the solution of systems of lineardiophantine equations.* Siam J. on Computing 11 (4) 687-708 (1982)

[**DKT**] Domich D., Kannan R., Trotter L. : *Hermite normal form computation using modulo determinant arithmetic*, Math. Oper. Res., 12, pp. 50-59 (1987)

[**Fru1**] Frumkin M.: *An application of modular arithmetic to the construction of algorithms solving systems of liear equations.* Soviet. Math. Dokl., 17, pp. 1165-1169. (1976)

[**Fru2**] Frumkin M.: *Polynomial time algorithms in the theory of linear diophantine equations.* In Fundamentals of computation theory, M. Karpinsky ed. LNCS 56. pp. 386-392. (1977)

[**Fru3**] Frumkin M.: *Complexity questions in number theory.* J. Soviet. Mat., 29, pp. 386-392 (1985)

[**HM**] Hafner J., McCurley K. : Asymptotically fast triangularization of matrices over rings.Siam J. Comput., 20 (6), 1068-1083 (1991)

[**Kap**] Kaplanski I. : *Elementary divisors and modules*. Transactions of the A.M.S., 66, (1949), 464-491.

[**KKS**] Kaltofen E., Krishnamoorthy M., Saunders B. D. : *Fast parallel computation of Hermite and Smith normal forms of polynomial matrices*. Siam J. on Algebraic and Discrete Methods 8 numéro 4 683-690 (1987)

[**Ili1**] Iliopoulos C. : *Worst-case complexity bounds on algorithms for computing the canonical structure of finite abelian groups and the Hermite and Smith normal forms of an integer matrice.* Siam J. on Computing 18 (4) 658-669 (1989)

[**Ili2**] Iliopoulos C. : *Worst-case complexity bounds on algorithms for computing the canonical structure of infinite abelian groups and solving systems of linear diophantine equations.* Siam J. on Computing 18 (4) 670-678 (1989)

[**KB**] Kannan R., Bachem A. : *Polynomial algorithms for computing the Smith and Hermite normal forms of an integer matrix.* Siam J. on Computing 8 (4) 499-507 (1979)

[**LLM**] Labhalla L., Lombardi H., Marlin R. : *Algorithmes de calcul de la réduction de Hermite d'une matrice à coefficients polynomiaux.* Soumis au colloque MEGA 92 (Nice)

[**MRR**] R. Mines, F. Richman, W. Ruitenburg : *A Course in Constructive Algebra* (Springer-Verlag; Universitext; 1988) .

[**Sam**] Samuelson P. A. : *A method for determining explicitly the coefficients of the characteristic equation* . Ann. Math. Stat. 13 (1942) 424-429.

[**Sch**] Schrijver A.: *Theory of integer and linear programming*. John Wiley. New-York. (1985)