Hermite Normal Form

Introduction & Basic Application

13 March 2007

Abstract

We will present the Hermite Normal Form for matrices with integral entries. All concepts, however, lift easily to the case of general Principal Ideal Domains instead of $\mathbb{Z}$. Finally, we apply our result to bases of lattices.

1 Introducing the Hermite Normal Form

Consider the set $\mathcal{M}_n$ of $n \times n$ matrices with integral entries and non-zero determinant. The subset of invertible matrices is a group and will be denoted by $\mathcal{GL}_n(\mathbb{Z})$. Its elements will be called *unimodular* in the sequel. $\mathcal{GL}_n(\mathbb{Z})$ acts on $\mathcal{M}_n$ by left multiplication, and we're interested in finding a class of representatives for the resulting orbits. Given a matrix $A \in \mathcal{M}_n$ the representative of the orbit that A belongs to is more commonly called a *normal form* for A. Imitating usual Gaussian elimination it's possible to transform the matrix A into an upper triangular matrix. We'll see that by choosing the diagonal elements to be non-negative, and by reducing the entries above the diagonal modulo the diagonal element, we get such a normal form. It's called the called the *Hermite Normal Form*.

$$\mathrm{HNF}_n := \left\{ H \in \mathcal{M}_n \colon \begin{array}{ll} (\forall i, j) & H_{ij} \geqslant 0 \\ (\forall i > j) & H_{ij} = 0 \\ (\forall i < j) & H_{ij} < H_{jj} \end{array} \right\}$$

The Hermite Normal Form has numerous applications for example in integer programming or cryptography.

Theorem 1. *Let $A \in \mathcal{M}_n$. Then there exists a unique $H \in \mathrm{HNF}_n$ such that $U\,A = H$ for some unimodular U.*

Proof.
"Existence". As indicated we'll first imitate Gaussian elimination using row operations. Let's start working on the first column. By interchanging rows we put the element a of least modulus at the $(1, 1)$ position. If a divides all the entries below it, we add multiples of the first row to make all these entries zero. If not there is an entry b below a such that b is not a multiple of a. So we find $q, r \in \mathbb{Z}$ such that

$$b = q\,a + r, \quad |r| < |a|.$$

By subtracting q times the first row, we hence get the element r in place of b. Since the modulus of r is smaller than the one of a, swap rows so that r is at $(1, 1)$ and repeat what we just did. Finally, all entries in the first column below the diagonal will be zero. Now move on to the second column...

Armin Straub
astraub@math.tulane.edu

"Uniqueness". If $U_1\,A = H_1$ and $U_2\,A = H_2$ where U_i are unimodular and $H_i \in \mathrm{HNF}_n$ then $(U_2\,U_1^{-1})\,H_1 = H_2$. Since $U_2\,U_1^{-1}$ is unimodular we see that it suffices to show the following: If $G, H \in \mathrm{HNF}_n$ and $U\,G = H$ for some unimodular U then $G = H$ or, equivalently, $U = I$. First, note that U has to be upper triangular as are G and H, since the i-th row of H written as a linear combination of rows of G can not include the first $i-1$ rows of G. Together with $\det U = \pm 1$ this implies that the diagonal entries of U have to be ± 1. In fact, they have to be 1 since G, H only have non-negative entries. Finally, suppose that $U \neq I$. Let i_0 be the index of the first row of U having a non-zero entry $U_{i_0\,j_0}$ where $j_0 > i_0$ is chosen to be minimal as well. Then

$$H_{i_0\,j_0} = \sum_{j=1}^{n} U_{i_0\,j}\,G_{j\,j_0} = G_{i_0\,j_0} + U_{i_0\,j_0}\,G_{j_0\,j_0}$$

using the triangularity of U and G. Hence $H_{i_0\,j_0} \equiv G_{i_0\,j_0} \bmod G_{j_0\,j_0}$. But $G_{j_0\,j_0} = H_{j_0\,j_0}$ since we already know that U has 1's on its diagonal. The condition on G and H that elements above the diagonal have to be smaller than the diagonal entry then implies $H_{i_0\,j_0} = G_{i_0\,j_0}$ contradicting the assumption that $U_{i_0\,j_0} \neq 0$. □

Remark 2. Starting with a matrix A we employed row operations (by left multiplication with unimodular matrices) to reduce it to its HNF which is upper triangular. If we allow column operations as well, we are able to even diagonalize A. With the further requirement that diagonal entries of the resulting matrix S are non-negative and divide each other, ie. $S_{i,i} | S_{i+1,i+1}$ we again obtain a normal form, called the *Smith Normal Form*. These two normal forms are of outstanding importance whenever it comes to efficient computation in relation to $\mathbb{Z}$-modules. They are of theoretic importance as well, as for instance usage of the SNF allows for an immediate proof of the Classification Theorem of finitely generated abelian groups.

2 Application to Lattices

A subgroup $\Gamma < \mathbb{R}^n$ is called a *lattice* if

$$\Gamma = \bigoplus \mathbb{Z}\,\omega_i$$

for a basis $\{\omega_i\}$ of $\mathbb{R}^n$. If $\Gamma' = \bigoplus \mathbb{Z}\,\alpha_i$ is a subgroup of Γ then every α_i is an integral linear combination of the ω_j. In other words,

$$\begin{pmatrix} — & \alpha_1 & — \\ & \vdots & \\ — & \alpha_n & — \end{pmatrix} = A \begin{pmatrix} — & \omega_1 & — \\ & \vdots & \\ — & \omega_n & — \end{pmatrix}$$

for some $A \in \mathbb{Z}^{n\times n}$. Every matrix $A \in \mathbb{Z}^{n\times n}$ corresponds in this fashion to exactly one subgroup, say $\Gamma(A)$, of Γ. Clearly,

$$\Gamma(A) = \Gamma(B) \quad \Longleftrightarrow \quad A = U\,B$$

for some unimodal U. Note that Γ' is a sublattice if and only if $\det A \neq 0$ since the α_i still have to form a basis of $\mathbb{R}^n$. These observations allow for the following immediate application of the Hermite Normal Form.

Corollary 3. *The map $H \to \Gamma(H)$ is a bijection between* HNF_n *and the sublattices of* Γ.

Note that we also have the following.

Proposition 4. *Let $A \in \mathcal{M}_n$. The index of $\Gamma(A)$ in Γ is given by $[\Gamma : \Gamma(A)] = |\det A|$.*

Proof. Let H be the HNF of A. Note that this leaves the modulus of the determinant unchanged. The triangularity of H now makes it easy to see that $[\Gamma : \Gamma(H)] = \det H$. □

Remark 5. Let Γ' be a sublattice of Γ of index m. Then $m\,\Gamma \leqslant \Gamma'$ and hence $m = [\Gamma : \Gamma'] = [\Gamma / m\,\Gamma : \Gamma' / m\,\Gamma]$. Note that $\Gamma / m\,\Gamma \cong \mathbb{Z}_m^n$ and that $\Gamma' / m\,\Gamma$ is a subgroup of order $m\,(n-1)$. This way the number of sublattices of index m can be seen to be equal to the number of subgroups of order $m\,(n-1)$ of $\mathbb{Z}_m^n$.

Armin Straub
astraub@math.tulane.edu

Finally, let's restate these results for lattices Γ in $\mathbb{C}$. Then $\Gamma = \mathbb{Z}\,\omega_1 \oplus \mathbb{Z}\,\omega_2$ for $\omega_1, \omega_2 \in \mathbb{C}$ linearly independent over $\mathbb{R}$.

Corollary 6. *The sublattices of Γ of index m are in bijective correspondence with the matrices*

$$\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in \mathbb{Z}^{2\times 2}_{\geqslant 0}$$

such that $a\,d = m$ and $b < d$.
In particular, the number of sublattices of index m is given by the sum of the divisors of m,

$$\sigma_1(m) = \sum_{d|m} d.$$

Example 7. Let p be prime. Then there are $p+1$ sublattices of Γ having index p. They correspond to the $p+1$ matrices

$$\begin{pmatrix} 1 & b \\ 0 & p \end{pmatrix} \text{ where } 0 \leqslant b < p, \quad \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}.$$

Armin Straub
astraub@math.tulane.edu