Quick Notes on Hermite Factorization

Hermite Factorization, which dates back to work of Charles Hermite around 1856, is to integer matrices what Gaussian Elimination is to matrices over fields. Hermite Factorization has applications to the classification of finitely generated Abelian groups and, more recently, quantum-resistant cryptography. Most relevant to our setting, Hermite Factorization simultaneously (a) generalizes the Extended Euclidean Algorithm, (b) enables the solution of binomial systems over any field, and (c) can be thought of as a (disguised) special case of the calculation of a Gröbner basis. Here, we'll very quickly review the core fundamentals of Hermite Factorization.

Recall that for any matrix $A = [a_{i,j}] \in \mathbb{C}^{m\times n}$, a *leading entry* of A is simply an entry $a_{i,j}$ with (i) $a_{i,j} \neq 0$, (ii) either $i = m$ or $a_{i',j} = 0$ for all $i' > i$, and (iii) either $j = 1$ or $a_{i,j'} = 0$ for all $j' < j$. For instance, if A happens to be in row echelon form, the leading entries of A are exactly the leading 1s. More concretely, the leading entries of

$$\begin{bmatrix} \boxed{-3} & 0 & 4 & 5 & 1 & 9 \\ 0 & 11 & -3 & 2 & 77 & -2 \\ 0 & \boxed{5} & 2 & 9 & -13 & 5 \\ 0 & 0 & 0 & \boxed{17} & 3 & 2 \end{bmatrix}$$

(which is not in row echelon form) are $\boxed{\text{boxed}}$ and lie in rows 1, 3, and 4.

Recall also that a square matrix U is called *unimodular* if and only if (a) all the entries of U are integers and (b) $\det U \in \{\pm 1\}$. A non-trivial example would be $U = \begin{bmatrix} 55 & 34 \\ 34 & 21 \end{bmatrix}$.

Definition 1 *Given any matrix $A \in \mathbb{Z}^{m\times n}$ we define a* Hermite Factorization of A *to be any identity of the form $UA = H$, where $U \in \mathbb{Z}^{m\times m}$ is unimodular, $H = [h_{i,j}] \in \mathbb{Z}^{m\times n}$ is upper-triangular, and:*

(a) every leading entry of H is positive, and

(b) for every leading entry $h_{i,j}$ of H we have $0 \leq h_{i',j} < h_{i,j}$ for all $i' < i$. $\diamond$

For instance, a Hermite factorization of

$$\begin{bmatrix} 77 & 10 & -7 & 2 \\ -5 & 44 & 2 & 18 \\ 9 & 2 & 3 & 5 \end{bmatrix} \text{ is } \begin{bmatrix} -19 & -3 & 161 \\ -32 & -5 & 271 \\ -203 & -32 & 1719 \end{bmatrix} \begin{bmatrix} 77 & 10 & -7 & 2 \\ -5 & 44 & 2 & 18 \\ 9 & 2 & 3 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 610 & 713 \\ 0 & 2 & 1027 & 1201 \\ 0 & 0 & 6514 & 7613 \end{bmatrix}.$$

Theorem 2 *Every matrix $A \in \mathbb{Z}^{m\times n}$ has a Hermite factorization. In particular, the factorization is unique if and only if A has rank m. In which case, we call the identity $UA = H$* the *Hermite factorization of A, and we call H* the Hermite normal form *of A.*

The hardest part of the theorem is actually the first statement, which asserts the existence of Hermite factorization. If one understands the case where A consists of a single column then the rest is easy.

Lemma 3 *For any $a_1, \ldots, a_m \in \mathbb{Z}$ we can always find a unimodular matrix $U \in \mathbb{Z}^{m\times m}$ with*

$$U \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \gcd(a_1, \ldots, a_m) \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Note: We use the standard convention that $\gcd(a_1, \ldots, a_m)$ is positive if it is nonzero.

Proof of Lemma 3: The case $m = 1$ is easy, since A is just the single entry a_1: If $a_1 = 0$ then we have $UA = H$ with $U = 1$ and $H = 0$. Otherwise, we have $UA = H$ with $U = \pm 1$ (where the sign agrees with that of a_1) and $H = |a_1|$.

The case $m=2$ is actually a disguised version of the Extended Euclidean Algorithm, applied to the two entries of A. From what we know about the Extended Euclidean Algorithm, we can always find a 2×2 unimodular matrix V with $[a_1, a_2]V = [\gcd(a_1, a_2), 0]$. Taking transposes, and setting $U = V^\top$, we are done.

For $m \geq 3$ we simply proceed by induction: Assume we can find we can find a unimodular $\bar{U} \in \mathbb{Z}^{(m-1)\times(m-1)}$ with $\bar{U} \begin{bmatrix} a_1 \\ \vdots \\ a_{m-1} \end{bmatrix} = \begin{bmatrix} \gcd(a_1, \ldots, a_{m-1}) \\ 0 \\ \vdots \\ 0 \end{bmatrix}$. Then, using block-multiplication, we also clearly obtain

$$\begin{bmatrix} \bar{U} & \\ & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \gcd(a_1, \ldots, a_{m-1}) \\ 0 \\ \vdots \\ 0 \\ a_m \end{bmatrix}.$$

So then we can easily find an elementary matrix S providing a swap so that

$$S \begin{bmatrix} \bar{U} & \\ & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \gcd(a_1, \ldots, a_{m-1}) \\ a_m \\ \vdots \\ 0 \end{bmatrix}.$$

Applying block multiplication and the $m=2$ case one last time, we can then find a 2×2 unimodular matrix W with $\begin{bmatrix} W & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} S \begin{bmatrix} \bar{U} & \mathbf{O} \\ \mathbf{O'} & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \gcd(a_1, \ldots, a_m) \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

So $U = \begin{bmatrix} W & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} S \begin{bmatrix} \bar{U} & \mathbf{O} \\ \mathbf{O'} & 1 \end{bmatrix}$ satisfies our desired factorization and we are done. ■

Lemma 4 *For any $a_1, \ldots, a_m \in \mathbb{Z}$ with $a_m \neq 0$ we can always find a unimodular matrix $U' \in \mathbb{Z}^{m\times m}$ with*

$$U' \begin{bmatrix} a_1 \\ \vdots \\ a_{m-1} \\ a_m \end{bmatrix} = \begin{bmatrix} a'_1 \\ \vdots \\ a'_{m-1} \\ |a_m| \end{bmatrix},$$

$a'_1, \ldots, a'_{m-1} \geq 0$, and $|a_m| > a'_1, \ldots, a'_{m-1}$.

Proof of Lemma 4: For each $i \in \{1, \ldots, m-1\}$ let a'_i be the unique element of $\{0, \ldots, |a_m|\}$ with $a_i \equiv a'_i \bmod a_d$. In particular, we can set $a'_i = a_i - \lfloor a_i/a_m \rfloor a_m$. Let I denote the $m \times m$ identity matrix. Then, letting E_i denote the $m \times m$ elementary matrix that equals I save for an entry of $-\lfloor a_i/a_m \rfloor$ in the (i, m) position (for all $i \in \{1, \ldots, m-1\}$), and letting E_m be the matrix obtained by replacing the (m, m) entry of I by $a_m/|a_m|$, we easily obtain $E_1 \cdots E_m \begin{bmatrix} a_1 \\ \vdots \\ a_{m-1} \\ a_m \end{bmatrix} = \begin{bmatrix} a'_1 \\ \vdots \\ a'_{m-1} \\ |a_m| \end{bmatrix}$. So we can take $U' := E_1 \cdots E_m$ and we are done. ■

Proof of Theorem 2: If A is the zero matrix then we can simply set U to be the $m \times m$ identity matrix and H to be the $m \times n$ zero matrix. Since there are no leading entries at all for H, we vacuously obtain a Hermite Factorization, and we see that the factorization is non-unique since we could have multiplied U by -1. So let us assume henceforth that A has at least one nonzero entry.

Recalling how Gaussian Elimination proceeds column by column to create leading 1s, we will mimic this process with elementary row operations *invertible over* $\mathbb{Z}$ to produce leading *entries* that are simply positive (and possibly not equal to 1).

In particular, by Lemma 3, we can find a matrix U_1 with $H_1 := U_1 A$ having a positive leading entry at the top of its first nonzero column. Let us call this entry h_{1,j_1}.

If $m=1$ or $j_1=n$ then we are done. Otherwise, we let A_2 be the submatrix of H_1 defined by rows $2,\ldots,m$ and columns $j_1+1,\ldots,n$ of H_1. By Lemma 3 once again, we can find an $(m-1)\times(m-1)$ unimodular matrix U_2 with $U_2 A_2$ having a leading entry at the top of its first nonzero column. Clearly, this implies that $\begin{bmatrix} 1 & \\ & U_2 \end{bmatrix}$ is unimodular and $\begin{bmatrix} 1 & \\ & U_2 \end{bmatrix} H_1$ has a leading entry in its second row, say at position j_2. By Lemma 4, we can then find a 2×2 matrix U_2' such that

$$H_2 := \begin{bmatrix} U_2' & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} \begin{bmatrix} 1 & \\ & U_2 \end{bmatrix} H_1 = \begin{bmatrix} 0 & \cdots & 0 & h_{1,j_1} & \cdots & h_{1,j_2-1} & h_{1,j_2} & * & \cdots & * \\ 0 & \cdots & 0 & 0 & \cdots & 0 & h_{2,j_2} & * & \cdots & * \\ \vdots & & \vdots & \vdots & & \vdots & 0 & * & \cdots & * \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & * & \cdots & * \end{bmatrix},$$

where $h_{2,j_2} > h_{1,j_2} \geq 0$ and * indicates an unknown entry. In particular, H_2 has leading entries in rows 1 and 2.

Continuing this way inductively, we arrive at the desired factorization, with $H = H_j$ for some $j \leq m$ and U a product of block-diagonal matrices with diagonal blocks of one of the following forms: 1, a unimodular matrix $U_i \in \mathbb{Z}^{(m-i+1)\times(m-i+1)}$, or a unimodular matrix $U_i' \in \mathbb{Z}^{i \times i}$, for $i \in \{1,\ldots,m-1\}$. This proves the first assertion of our theorem, on the existence of Hermite factorization.

To prove the final assertion, first assume that A has rank m.

Then every Hermite factorization of A must result in an H with exactly m leading entries. Otherwise, we could divide each row with a leading entry by its leading entry and then use Gaussian Elimination to obtain a row echelon form for A with strictly less than m leading 1s, implying that A has rank $< m$ — a contradiction. So let $U_1 A = H_1$ and $U_2 A = H_2$ be any Hermite factorizations for A. We will show that $U_1 = U_2$ and $H_1 = H_2$.

Since U_1 and U_2 are unimodular, they are invertible, and we immediately obtain that $A = U_1^{-1} H_1 = U_2^{-1} H_2$ and thus $U_2 U_1^{-1} H_1 = H_2$. Letting $U' = U_2 U_1^{-1}$ it is clear that U' is also unimodular (and invertible). Therefore, H_1 and H_2 have the same number of zero columns.

For $i \in \{1,2\}$ let H_i' denote the $m \times m$ (upper-triangular and invertible) sub-matrix of H_i consisting of the columns of H_i containing leading entries. Since upper-triangular matrices are closed under product and inverse, $U' = H_2' H_1'^{-1}$ is upper-triangular as well. Therefore, the diagonal entries of U' have absolute value 1. We will in fact ultimately prove that U' is the $m \times m$ identity matrix, and thus $U_1 = U_2$ and $H_1 = H_2$.

Toward this end, let $[u'_{i,j}]_{m\times m} := U'$, $[h'(1)_{i,j}]_{m\times m} := H_1'$, and $[h'(2)_{i,j}]_{m\times m} := H_2'$. Since $h'(1)_{1,1}, h'(2)_{1,1}, \ldots, h'(1)_{m,m}, h'(2)_{m,m} > 0$, and $u'_{i,i} h'(1)_{i,i} = h'(2)_{i,i}$ for all i (since u' is upper-triangular), we see that the diagonal entries of U' are in fact all 1. Moreover, we see that the *diagonal* entries of H_1' and H_2' must be identical.

Now observe that $u'_{1,1} h'(1)_{1,2} + u'_{1,2} h'(1)_{2,2} = h'(2)_{1,2}$, and thus $h'(1)_{1,2} + u'_{1,2} h'(1)_{2,2} = h'(2)_{1,2}$. In particular, if $u'_{1,2} > 0$ then $h'(2)_{1,2} > h'(1)_{2,2} = h'(2)_{2,2}$ (since $h'(1)_{1,2} > 0$) and

this contradicts the definition of Hermite factorization (since we must have $h'(2)_{1,2} < h'(2)_{2,2}$. Similarly, if $u'_{1,2} < 0$ then we obtain $h'(1)_{1,2} > h'(1)_{2,2}$ again yielding a contradiction. So we must have $u'_{1,2} = 0$.

Repeating the last argument inductively, computing $h'(2)_{1,3}, \ldots, h'(2)_{1,m}$, we successively obtain that $u'_{1,3} = 0, \ldots, u'_{1,m} = 0$. So the first row of U' agrees with the first row of the $m \times m$ identity matrix. Continuing in the same way with the second row, the third row, etc., we then at last obtain U' is the $m \times m$ identity matrix. So $U_1 = U_2$ and thus $H_1 = H_2$.

To conclude, assume that the rank of A is strictly less than m. We need to show that there at least two different Hermite factorizations for A. This is easy, upon observing that our low rank assumption implies that there is at least one row of H — call it h_i — with no leading entry. Should this $h_i = \mathbf{O}$ then we can clearly add any integral multiple of h_i to any other row of H leaving H unchanged. This means that there are infinitely many integral shears that we can pre-multiply U by to get the same H, so U was not unique.

If h_i is not the zero row then it must have a positive non-zero entry $h_{i,j}$ with $j \geq 2$ and $h_{i,j-1} = 0$. So let $h_{i'}$ be the unique row of H with a leading entry in the $j\underline{\text{th}}$ column. (Note that $i' > i$ since H is upper-triangular.) Then we can add any integral multiple of h_i to $h_{i'}$ and the resulting new H (and the resulting new U obtained by pre-multiplying the old U by a suitable shear) will still satisfy all the axioms of Hermite factorization. So neither U nor H are unique. ■