Let $\widehat{G} = \{M \in GL_n(\mathbb{Z}) \mid M \text{ is upper triangular}\}$. The main diagonal of any matrix of this form must consist entirely of 1's and $-1$'s, so there is a surjection $\phi\colon \widehat{G} \twoheadrightarrow \{\pm 1\}^n \cong (\mathbb{Z}/2)^n$ given by forgetting everything except the diagonal. The kernel $G$ is the subgroup of unitriangular matrices over $\mathbb{Z}$. There is an obvious section of $\phi$, so $\widehat{G}$ splits as a semi-direct product: $\widehat{G} = G \rtimes (\mathbb{Z}/2)^n$.

Now define:

$$\Upsilon = \bigl\{ r = (r_1, r_2, r_3, \ldots) \bigm| r_i \in \mathbb{Z},\, r_i \geqslant 0, \text{ each } r_i \text{ divides } r_{i-1} \bigr\}.$$

This set has an operation $\Upsilon \times \Upsilon \longrightarrow \Upsilon$ defined by $(r,s) \mapsto r \oplus s = t$, where we set $t_1 = 0$ and $t_i = \gcd(\{r_j s_k \mid j+k=i\})$ for $i \geqslant 2$. This turns $\Upsilon$ into an abelian semigroup. There is also a partial order on $\Upsilon$, defined by $r \lessdot s$ if and only if $r_i$ divides $s_i$ for all $i \geqslant 1$. There is a smallest element $\langle 1 \rangle = (1,1,1,\ldots)$ for this partial order and a largest element $\langle \infty \rangle = (0,0,0,\ldots)$. In general, we write $\langle k \rangle = (0,0,\ldots,0,1,1,\ldots) \in \Upsilon$, where there are $k-1$ zeros, for $k \in \{1,2,3,\ldots,\infty\}$. Note that $\langle k \rangle \oplus \langle l \rangle = \langle k+l \rangle$. Any two elements of $\Upsilon$ have a unique infimum and supremum (which we write as $\mathsf{min}$ and $\mathsf{max}$ respectively), so $(\Upsilon, \lessdot)$ is a bounded lattice. The structures $\oplus$ and $\lessdot$ on $\Upsilon$ are compatible in the sense that $r \lessdot s$ implies that $r \oplus t \lessdot s \oplus t$ and $t \oplus r \lessdot t \oplus s$. Hence $(\Upsilon, \oplus, \lessdot)$ is a so-called *abelian posemigroup*. We also note that $\langle k \rangle \oplus r = (0,\ldots,0,r_1,r_2,\ldots) \gtrdot r$ and $r \oplus r \gtrdot r$.

For $k \geqslant 0$ let $U_k$ be the set of $(n \times n)$-matrices $\{a_{ij}\}$ where $a_{ij} = 0$ if $j < i+k$. So $U_0$ is the set of upper triangular matrices (which are not necessarily invertible) and $U_k = \{0\}$ for $k \geqslant n$. For $r \in \Upsilon$ define

$$U_r = \bigl\{ r_1 A_1 + r_2 A_2 + \cdots + r_{n-1} A_{n-1} \bigm| A_k \in U_k \bigr\},$$

so $U_k = U_{\langle k \rangle}$. The key properties of this definition (not too hard to check) are that for any $r, s \in \Upsilon$,

$$X \in U_r \text{ and } Y \in U_s \quad \Rightarrow \quad X+Y \in U_{\mathsf{min}(r,s)} \text{ and } XY \in U_{r \oplus s}, \tag{1}$$

$$r \lessdot s \text{ in } \Upsilon \quad \Leftrightarrow \quad U_r \supseteq U_s. \tag{2}$$

In particular, every $U_r$ is contained in $U_{\langle 1 \rangle}$, which is the set of strictly upper triangular matrices. Next, we define

$$H_r = \{ I + A \mid A \in U_r \} \subseteq \widehat{G}.$$

It's not too hard to see that $H_r$ is a subgroup of $\widehat{G}$. There are generating sets for $H_r$ and for $\widehat{G}$ as follows. Let $E_{i,j}$ be the matrix whose entries are all 0 except for the $(i,j)$-th one, which is 1. Let $J_i$ be the diagonal matrix whose diagonal entries are all 1, except for the $(i,i)$-th one, which is $-1$. Then $H_r$ is generated by the set

$$\bigl\{ I + r_k E_{i,i+k} \bigm| k \in \{1,\ldots,n-1\} \text{ and } i \in \{1,\ldots,n-k\} \bigr\}$$

and $\widehat{G}$ is generated by the set

$$\bigl\{ I + E_{i,j} \bigm| i,j \in \{1,\ldots,n\}, i<j \bigr\} \cup \bigl\{ J_i \bigm| i \in \{1,\ldots,n\} \bigr\}.$$

Some useful commutator identities that these basic matrices satisfy are as follows. For $i<j$ and $k<l$ we have:

$$\begin{aligned} [I + rE_{i,j}, I + sE_{k,l}] &= \begin{cases} I & j \neq k \text{ and } i \neq l \\ I + rsE_{i,l} & j = k \\ I - rsE_{k,j} & i = l \end{cases} \\ [I + E_{i,j}, J_k] &= \begin{cases} I & i,j,k \text{ pairwise distinct} \\ I + 2E_{i,j} & k = i \text{ or } k = j \end{cases} \end{aligned} \tag{3}$$

and $[J_i, J_j] = I$ for any $i,j$.

Now, for any $(n \times n)$-matrix $X$, if $X$ is nilpotent, i.e. $X^\ell = 0$ for some $\ell$, then $I+X$ is invertible and we define $\overline{X} = (I+X)^{-1}$. This satisfies the relation $\overline{X} = I - X\overline{X}$, and hence $X - X\overline{X} - XX\overline{X} = 0$. Using this, we see that, for any matrix $Y$,

$$(I+X)(I+Y)(I+X)^{-1} = I + Y + XY - YX\overline{X} - XYX\overline{X}.$$

A corollary of this is that, for any $r \in \Upsilon$, the subgroup $H_r \leqslant \widehat{G}$ is *normal*. To see this, let $D$ be a diagonal matrix with all diagonal entries $\pm 1$, let $X \in U_{\langle 1 \rangle}$ and let $Y \in U_r$. We need to show that

$$D(I+X)(I+Y)(I+X)^{-1}D^{-1}$$

is of the form $I + Z$ for some $Z \in U_r$. Since $X$ and $Y$ are strictly upper triangular, they are nilpotent, so we may apply the equation above to see that this is equal to

$$I + D(Y + XY - YX\overline{X} - XYX\overline{X})D^{-1}.$$

Conjugating a matrix by $D$ only changes the signs of some of the entries, so it has no effect on whether the matrix is contained in $U_r$ or not. So we can ignore this, and we just need to show that $Y + XY - YX\overline{X} - XYX\overline{X}$ is in $U_r$. Obviously $Y \in U_r$, by assumption. By property (1), $XY \in U_{\langle 1\rangle \oplus r}$, which is contained in $U_r$ by property (2), since $\langle 1\rangle \oplus r \geqslant r$. The matrix $\overline{X}$ is upper triangular, and $X$ is strictly upper triangular, so their product is also strictly upper triangular, i.e. $X\overline{X} \in U_{\langle 1\rangle}$. So by properties (1) and (2), $YX\overline{X} \in U_{\langle 1\rangle \oplus r} \subseteq U_r$ and $XYX\overline{X} \in U_{\langle 2\rangle \oplus r} \subseteq U_r$. Finally, by property (1) again, the sum $Y + XY - YX\overline{X} - XYX\overline{X}$ is in $U_r$, which completes the proof that $H_r$ is normal in $\widehat{G}$.

In general, suppose that $\Gamma$ is a group, $S \subseteq \Gamma$ is a generating set for $\Gamma$ and $N \lhd \Gamma$ is a normal subgroup, such that $[s,t] \in N$ for all $s,t \in S$. Then $[\Gamma,\Gamma] \leqslant N$. This can be proved using the following commutator identities, which hold for any $x,y,z$ in any group:

$$[x,zy] = [x,y](y^{-1}[x,z]y) \qquad [x,y^{-1}] = y[y,x]y^{-1}$$
$$[xz,y] = (z^{-1}[x,y]z)[z,y] \qquad [x^{-1},y] = x[y,x]x^{-1}$$

Now we apply this fact, with $\Gamma = \widehat{G}$, with the generating set described above, and $N = H_{(2,1,1,\ldots)}$. Using the commutator identities (3), we see that every commutator $[s,t]$, with $s$ and $t$ in the generating set for $\widehat{G}$, lies in the normal subgroup $H_{(2,1,1,\ldots)}$. Thus, from the previous paragraph, we deduce that $[\widehat{G},\widehat{G}] \leqslant H_{(2,1,1,\ldots)}$. On the other hand, the commutator identities (3) also show that every generator of $H_{(2,1,1,\ldots)}$ can be written as a commutator of elements of $\widehat{G}$. Thus we have:

$$[\widehat{G},\widehat{G}] \;=\; H_{(2,1,1,\ldots)}. \tag{4}$$

Now fix any element $r \in \Upsilon$ and set $\Gamma = H_r$, with its generating set described above. Since $r \oplus r \geqslant r$ we have $H_{r\oplus r} \leqslant H_r \leqslant \widehat{G}$. Since $H_{r\oplus r}$ is normal in $\widehat{G}$, it is also normal in $H_r$, so we may take $N = H_{r\oplus r}$. Using the commutator identities (3), as before, we can check that (a) every commutator $[s,t]$ of generators $s$ and $t$ of $H_r$ is contained in $H_{r\oplus r}$, and conversely (b) every generator of $H_{r\oplus r}$ may be written as a commutator of elements of $H_r$. Thus we have:

$$[H_r,H_r] \;=\; H_{r\oplus r}. \tag{5}$$

Indeed, the operation $\oplus\colon \Upsilon \times \Upsilon \longrightarrow \Upsilon$ was defined exactly in order to make this fact true, and more generally:

$$[H_r,H_s] \;=\; H_{r\oplus s}, \tag{6}$$

for any two elements $r,s \in \Upsilon$.

Let's define a sequence $[i]$ of elements of $\Upsilon$ by $[1] = (2,1,1,1,\ldots)$ and $[i+1] = [i] \oplus [i]$ for $i \geqslant 1$. Then:

$$[i+1] \;=\; (0,0,\ldots,0,0,2^{2^i},2^{2^i-1},2^{2^i-2},\ldots,4,2,1,1,1,\ldots),$$

where there are $2^i - 1$ zeros. Facts (4) and (5) imply inductively that:

$$\widehat{G}^{(i)} \;=\; H_{[i]},$$

so we have an explicit description of the derived series of $\widehat{G}$. Note that $H_{[i]} = \{0\}$ if and only if $2^{i-1} \geqslant n$, so the derived length of $\widehat{G}$ is $\lceil \log_2(n) \rceil + 1$.

For comparison, the derived series of the subgroup $G \leqslant \widehat{G}$ is $G^{(i)} = H_{\langle 2^i\rangle} \leqslant H_{[i]} = \widehat{G}^{(i)}$, see here, and its derived length is exactly one less than the derived length of $\widehat{G}$.

Acknowledgement: *Thanks to Jonas Antor for pointing out to me* (a) *that my initial "solution" for the derived series of* $\widehat{G}$ *was wrong – I had instead computed the derived series of its index-*$2^n$ *subgroup* $G$*, see also here – and moreover* (b) *fact* (4) *above.*