Smith Normal Form

We say that \( A \in F[t]^{m \times n} \) is in **Smith Normal Form** (or **SNF**) if \( A \) is the zero matrix or if \( A \) is a matrix of the form

\[
\begin{pmatrix}
q_1 & 0 & \cdots \\
0 & q_2 & \\
\vdots & & \ddots \\
& & & q_r \\
0 & & & & 0 & \cdots \\
\end{pmatrix}
\]

with \( q_1 | q_2 | q_3 | \cdots | q_r \) in \( F[t] \) and all monic, i.e., there exists a positive integer \( r \) satisfying \( r \leq \min(m, n) \) and \( q_1 | q_2 | q_3 | \cdots | q_r \) monic in \( F[t] \) such that \( A_{ii} = q_i \) for \( 1 \leq i \leq r \) and \( A_{ij} = 0 \) otherwise.

We generalize Gaussian elimination, i.e., row (and column) reduction for matrices with entries in \( F \) to matrices with entries in \( F[t] \). The only difference arises because most elements of \( F[t] \) do not have multiplicative inverses.

Let \( A \in M_n(F[t]) \). We say that \( A \) is an **elementary matrix** of

i). **Type I:** if there exists \( \lambda \in F[t] \) and \( l \neq k \) such that

\[
A_{ij} = \begin{cases} 
1 & \text{if } i = j \\
\lambda & \text{if } (i, j) = (k, l) \\
0 & \text{otherwise}
\end{cases}
\]

ii). **Type II:** if there exists \( k \neq l \) such that

\[
A_{ij} = \begin{cases} 
1 & \text{if } i = j \neq l \text{ or } i = j \neq k \\
0 & \text{if } i = j = l \text{ or } i = j = k \\
1 & \text{if } (k, l) = (i, j) \text{ or } (k, l) = (j, i) \\
0 & \text{otherwise}
\end{cases}
\]

iii). **Type III:** if there exists a \( 0 \neq u \in F[t] \) and \( l \) such that

\[
A_{ij} = \begin{cases} 
1 & \text{if } i = j \neq l \\
u & \text{if } i = j = l \\
0 & \text{otherwise}
\end{cases}
\]

**Remarks.** Let \( A \in F[t]^{m \times n} \). Multiplying \( A \) on the left (respectively right) by a suitable size elementary matrix of

a.) **Type I** is equivalent to adding a multiple of a row (respectively column) of \( A \) to another row (respectively column) of \( A \).

b.) **Type II** is equivalent to interchanging two rows (respectively columns) of \( A \).
c). Type III is equivalent to multiplying a row (respectively column) of $A$ by an an element in $F[t]$ having a multiplicative inverse.

**Remarks.**
1). All elementary matrices are invertible.
2). The definition of elementary matrices of Types I and II is exactly the same as that given when defined over a field.
3). The elementary matrices of Type III have a restriction. The $u$’s appearing in the definition are precisely the elements in $F[t]$ having a multiplicative inverse. The reason for this is so that the elementary matrices of Type III are invertible.

**Notation.** We let

\[
\text{GL}_n(F[t]) := \{ A \mid A \text{ is invertible} \}.
\]

**Warning:** a matrix in $M_n(F[t])$ having $\det(A) \neq 0$ may no longer be invertible, i.e., have an inverse. What is true is that $\text{GL}_n(F[t])$ consists of those matrices whose determinant have a multiplicative inverse in $F[t]$.

**Definition.** Let $A, B \in F[t]^{m \times n}$. We say that $A$ is equivalent to $B$ and write $A \approx B$ if there exist matrices $P \in \text{GL}_m(F[t])$ and $Q \in \text{GL}_n(F[t])$ such that $B = P A Q$.

**Theorem.** Let $A \in F[t]^{m \times n}$. Then $A$ is equivalent to a matrix in Smith Normal Form (SNF). Moreover, there exist matrices $P \in \text{GL}_m(F[t])$ and $Q \in \text{GL}_n(F[t])$, each a product of matrices of Type I, Type II, and Type III, such that $P A Q$ is in SNF.

**Proof.** The proof will, in fact, be an algorithm to find a SNF for $A$. Let $\deg(f)$ denote the degree of a polynomial $0 \neq f \in F[t]$.

If $A = 0$ there is nothing to do, so assume that $A \neq 0$. Let $A = (a_{ij})$.

**Step 1.** Choose $a = a_{ij} \neq 0$ such that $\deg(a)$ is minimal among all the $\deg(a_{lk})$, $a_{lk} \neq 0$. Put $a$ in the $(1,1)$ spot using matrices of Type II. In particular, we may assume that $a = a_{11}$. In addition, use a Type III matrix to make $a$ monic.

**Step 2.** If $a \mid a_{ij}$ in $F[t]$ for all $i$ and $j$, use Type I matrices to transform $A$ into a matrix of the form

\[
\begin{pmatrix}
a & 0 & \ldots & 0 \\
0 & & & \\
\vdots & & A_1 & \\
0 & & & \\
\end{pmatrix}
\]

In this case, $a$ divides every entry of $A_1$ [check] and every non-zero entry of $A_1$ has its degree $\geq \deg(a)$. In particular, $\deg a$ is still minimal. Now take the matrix $A_1$ and go back to Step 1.

[Note. If this occurs, by induction there exist invertible matrices $Q_1, P_1$, each a product of elementary matrices, such that $P_1 A_1 Q_1$ is in SNF. Let]
\[ P = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & & 0 \\ \vdots & & \ddots & P_1 \\ 0 & & & 0 \end{pmatrix} \quad \text{and} \quad Q = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & & 0 \\ \vdots & & \ddots & Q_1 \\ 0 & & & 0 \end{pmatrix} \]
then
\[
P \begin{pmatrix} a & 0 & \cdots & 0 \\ 0 & \ddots & & 0 \\ \vdots & & \ddots & A_1 \\ 0 & \cdots & & 0 \end{pmatrix} Q
\]
is in SNF and the Theorem easily follows.]

**Step 3.** Step 2 does not apply and there exists an entry \( b = a_{ij} \) in either the first row or first column such that \( a \nmid b \):
Using the Division Algorithm, write \( b = qa + r \) in \( F[t] \) with \( r \neq 0 \) and \( \deg(r) < \deg(a) \).
Use Type I matrices to change \( A \) into a matrix with \( r \) in it. Since \( \deg(a) \) is not longer minimal, go back to Step 1.
[Since \( \deg(a) \) is a non-negative integer and \( \deg(r) < \deg(a) \), this must eventually stop.]

**Step 4.** Neither Step 2 nor Step 3 apply. Thus \( a \mid a_{ij} \) whenever \( i = 1 \) or \( j = 1 \):
Use Type I matrices to convert \( A \) to
\[
\begin{pmatrix} a & 0 & \cdots & 0 \\ 0 & \ddots & & 0 \\ \vdots & & \ddots & A_1 \\ 0 & \cdots & & 0 \end{pmatrix}
\]
Now one of the following occurs.
(a) There exists a non-zero entry \( b \) in \( A_1 \) such that \( \deg(b) < \deg(a) \). So \( \deg(a) \) is no longer minimal. Go back to Step 1.
(b) \( a \mid b \) for all entries \( b \) in \( A_1 \). This is impossible — You should have been in Step 2.
[No matter, take \( A_1 \) and go to Step 1.]
(c) There exists an entry \( b \) in \( A_1 \) such that \( a \nmid b \):
Using the Division Algorithm, write \( b = qa + r \) in \( F[t] \) with \( r \neq 0 \) and \( \deg(r) < \deg(a) \).
Use Type II matrices to get \( b \) into the first column. (This does not change the \((1,1)\) entry \( a \).) Now use Type I matrices to change \( b \) to \( r \). Since \( \deg(a) \) is no longer minimal, go back to Step 1.
Clearly this algorithm yields a SNF of \( A \). \( \square \)

**Remark.** The SNF derived by this algorithm is, in fact, unique. In particular the monic polynomials \( q_1 \mid q_2 \mid q_3 \mid \cdots \mid q_r \) arising in the Smith Normal Form of a matrix \( A \) are unique and are called the **invariant factors** of \( A \). This is proven using results about determinants.
It follows if \( A, B \in \mathbb{F}^{m \times n} \) then \( A \approx B \) if and only if they have the same SNF if and only if they have the same invariant factors.

So what good is the SNF relative to linear operators on finite dimensional vector spaces? It tells us a great deal, because the following is true: Let \( A, B \in \mathbb{M}_n(\mathbb{F}) \). Then \( A \sim B \) (i.e., \( A \) and \( B \) are similar) if and only if \( tI - A \approx tI - B \) in \( \mathbb{M}_n(\mathbb{F}[t]) \) (i.e., \( tI - A \) and \( tI - B \) are equivalent) and this is completely determined by the SNF hence the invariant factors of \( tI - A \) and \( tI - B \). Now the SNF of \( tI - A \) may have some of its invariant factors 1, and we shall drop these (cf. below).

**Definition.** Let \( q = t^n + a_{n-1}t^{n-1} + \cdots + a_1t + a_0 \) be a monic polynomial in \( \mathbb{F}[t] \). The **companion matrix** \( C(q) \) of \( q \) is defined to be the \( n \times n \) matrix:

\[
\begin{pmatrix}
0 & 0 & \cdots & 0 & -a_0 \\
1 & 0 & \cdots & 0 & -a_1 \\
\vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & 1 & -a_{n-1}
\end{pmatrix}
\]

Let \( V \) be a finite dimensional vector space over \( \mathbb{F} \) with \( \mathcal{B} \) an ordered basis. Let \( T : V \to V \) be a linear operator. If one computes the Smith Normal Form of \( tI - [T]_\mathcal{B} \), it will have the form

\[
\begin{pmatrix}
1 & 0 & \cdots & \cdots & 0 \\
0 & 1 & \cdots & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & \cdots & q_r \\
0 & 0 & \cdots & \cdots & q_1
\end{pmatrix}
\]

with \( q_1 \mid q_2 \mid \cdots \mid q_r \) are all the monic polynomials in \( \mathbb{F}[t] \setminus \mathbb{F} \). These are called the **invariant factors** of \( T \). They are uniquely determined by \( T \). The main theorem is that there exists an ordered basis \( \mathcal{B} \) for \( V \) such that

\[
[T]_\mathcal{B} = \begin{pmatrix}
C(q_1) & 0 & \cdots & 0 \\
0 & C(q_2) & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & C(q_r)
\end{pmatrix}
\]

and this matrix representation is unique. This is called the **rational canonical form** or **RCF** of \( T \). Moreover, the minimal polynomial \( q_T \) of \( T \) is \( q_r \). The algorithm computes this as well as all invariant factors of \( T \). The characteristic polynomial \( f_T \) of \( T \) is the product \( q_1 \cdots q_r \). This works over any field \( \mathbb{F} \), even if \( q_T \) does not split. The basis \( \mathcal{B} \) gives a
decomposition of $V$ into $T$-invariant subspaces $V = W_1 \oplus \cdots \oplus W_r$ where $f_{T|_{W_i}} = q_{T|_{W_i}} = q_i$ and if $\dim(W_i) = n_i$ then $B_i = \{v_i, Tv_i, \ldots, T^{n_i-1}v_i\}$ is a basis for $W_i$. (We say that the $W_i$ are $T$-cyclic subspaces.)

Let $V$ be a finite dimensional vector space over $F$ with $B$ an ordered basis. Let $T : V \to V$ be a linear operator. Suppose that $q_T$ splits over $F$. Then we know that there exists a Jordan canonical form (JCF) of $T$. How do we compute it? We use the Smith normal form of $tI - [T]_B$ to compute the invariant factors $q_1 | q_1 | \cdots | q_r$ of $T$ just as one does to compute the RCF of $T$. We then factor each $q_i$. Suppose this factorization is

$$q_i = (t - \lambda_1)^{r_1} \cdots (t - \lambda_m)^{r_m}$$

in $F[t]$, with $\lambda_1, \ldots, \lambda_m$ distinct.

[Note that $q_{i+1}$ has this as a factor so it has the form

$$q_{i+1} = (t - \lambda_1)^{s_1} \cdots (t - \lambda_m)^{s_m} \cdots (t - \lambda_{m+k})^{s_{m+k}}$$

for some $k \geq 0$ with $s_i \geq r_i$ for each $1 \leq i \leq m$ and $s_{m+1}, \ldots, s_{m+k} \geq 1$ with $\lambda_1, \ldots, \lambda_{m+k}$ distinct.]

Then the totality of all the $(t - \lambda_i)^{r_j}$, including repetitions if they occur in different $q_i$’s, give all the elementary divisors of $T$. So to get the JCF of $T$ we take for each $q_i$ as factored above the block matrix

$$\begin{pmatrix}
J_{r_1}(\lambda_1) & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & J_{r_m}(\lambda_m)
\end{pmatrix}$$

and replace $C(q_i)$ by it in the RCF, i.e., we take all the Jordan blocks $J_r(\lambda)$ associated to each and every factor of the form $(t - \lambda)^{r_j}$ in each and every invariant factor $q_i$ determined by the SNF and form a matrix out of all such blocks. This is the JCF (which is unique only up to block order).

[Note we did not order the blocks having the same eigenvalue. It is nicer to do this also.]