Theorem 0.1. Let $S$ and $T$ be possibly unbounded operators acting in the Hilbertspace $H$. Then

$$
\begin{equation*}
T^{*} S^{*} \subset(S T)^{*} \tag{0.1}
\end{equation*}
$$

Furthermore, if in addition $S \in B(H)$ then

$$
\begin{equation*}
T^{*} S^{*}=(S T)^{*} \tag{0.2}
\end{equation*}
$$

Since the proof of this theorem is based on the definition of the adjoint of a given aperator, we start this proof by recalling this definition. It is given in equations (13.2) and (13.3) of Rudin. An equivalent formulation is the following:

Definition 0.2. Let $A$ be a given densely defined, possibly unbounded operator. Then its adjoint, $A^{*}$ is defined by,

$$
\begin{equation*}
D\left(A^{*}\right)=\left\{y: \text { there is a } y^{*} \text { such that for every } x \in D(A),(A x, y)=\left(x, y^{*}\right)\right\} \tag{0.3}
\end{equation*}
$$

and

$$
\begin{equation*}
A^{*} y=y^{*} \tag{0.4}
\end{equation*}
$$

To prove of conclusion (0.1) assume that

$$
\begin{equation*}
y \in D\left(T^{*}\right) S^{*} \tag{0.5}
\end{equation*}
$$

Then, we see from assumption (0.5) that $y \in D\left(S^{*}\right)$. This fact allows us to apply the definitions (0.3) and (0.4) to the operator $S$ in place of the operator $A$ and to the vector $T x$ in place of the vector $x$. This application yields,

$$
(S T x, y)=\left(T x, S^{*} y\right), \quad x \in D(S T)
$$

We also see from assumption $(0.5)$ that $S^{*} y \in T\left(S^{*}\right.$. This fact allows us to apply the definitions $(0.3)$ and (0.4) to the operator $T$ in place of the operator $A$ and to the vector ${ }^{*} y$ in place of the vector $y$. This application yields,

$$
\left.\left(T x, S^{*} y\right)=\left(x, T^{*} S^{*} y\right)\right), \quad x \in D(S T)
$$

Combining the previous to formulae we find

$$
\begin{equation*}
\left.(S T x, y)=\left(x, T^{*} S^{*} y\right)\right), x \in D(S T) \tag{0.6}
\end{equation*}
$$

On the other hand, applying the definitions $(0.3)$ and $(0.4)$ to the operator $S T$ in place of the operator $A$, we find

$$
\begin{equation*}
\left.(S T x, y)=\left(x, T^{*} S^{*} y\right)\right), \quad x \in D(S T) \tag{0.7}
\end{equation*}
$$

Finally, combining formulae $(0.3)$ and $(0.4)$ that these operators are densely we arrive at conclusion $(0.1)$.
Having established conclusion (0.1), to prove conclusion (0.2) all we have to do, is to show that

$$
\begin{equation*}
y \in D\left((S T)^{*}\right) \text { implies } y \in D\left(T^{*} S^{*}\right) \tag{0.8}
\end{equation*}
$$

Since by assumption $S \in B(H)$, we see from Section 12.9 that this also holds for the adjoint, $S^{*} \in B(H)$. Hence, $D\left(S^{*}\right)=H$ and so, we can prove this implication by proving that

$$
\begin{equation*}
S^{*} y \in D\left(T^{*}\right) \tag{0.9}
\end{equation*}
$$

Applying the definition (0.3) to the operator $T$ in place of the operator $A$, and to the vector $S^{*} y$ in place of the vecrtor $y$, we see that the implication (0.6) is equivalent to:

$$
\begin{equation*}
y \in D\left((S T)^{*}\right) \text { implies that there is a } y^{*} \text { such that for every } x \in D((T)),\left(T x, s^{*} y\right)=\left(x, y^{*}\right) \tag{0.10}
\end{equation*}
$$

Indeed, we claim that for $y \in D\left((T S)^{*}\right)$ the vector given by,

$$
\begin{equation*}
y^{*}=(T S)^{*} y \tag{0.11}
\end{equation*}
$$

is such a vector. To see this, first we apply the definition (0.3) to the operator $S$ in place of the operator $A$ and to the vector $T x$ in plac of the vector $x$. This application yields,

$$
(S T x, y)=\left(T x, S^{*} y\right)
$$

Second, we to apply the the definitions $(0.3)$ and $(0.2)$ to the operator $S T$ in place of the operator $A$. This application yields,

$$
(S T x, y)=\left(x,(S T)^{*} y\right), \quad x \in D(S T)
$$

Combining the previous two formulae and using the definition (0.11), we find

$$
\begin{equation*}
\left(T x, S^{*} y\right)=\left(x, y^{*}\right), x \in D(S T) \tag{0.12}
\end{equation*}
$$

Using the assumption $S \in B(H)$ again, we see that $D(S T)=D(T)$. Combining this fact with formula (0.12) we find the implication (0.10). Hence relation (0.9) follows and so does the implication (0.8). This completes the proof of conclusion (0.2).

