Full rank case. Rank$(X) = p$. Since the columns of X are linearly independent, there exists a unique vector $\hat{\beta}$ such that $\hat{\theta} = X\hat{\beta}$. (If the columns of X are not linearly independent $\hat{\beta}$ is not unique.) Therefore,

$$X'X\hat{\beta} = X'Y.$$

Since X has full column rank, $X'X$ is nonsingular. Therefore,

$$\hat{\beta}_{LS} = (X'X)^{-1}X'Y$$

is unique. One could also use calculus for this derivation:

$$||Y - X\beta||^2 = (Y - X\beta)'(Y - X\beta) = Y'Y - 2\beta'X'Y + \beta'X'X\beta,$$

so differentiating it w.r.t. β:

$$-2X'Y + 2X'X\beta = 0, \text{ or } X'X\hat{\beta} = X'Y.$$

Note that

$$\hat{\theta} = X\hat{\beta} = X(X'X)^{-1}X'Y = PY = \hat{Y},$$

where P is the projection matrix onto $\mathcal{M}_C(X)$.
$\hat{\epsilon} = Y - \hat{Y} = Y - X\hat{\beta} = (I - P)Y =$ residuals.

$$\begin{aligned}
\hat{\epsilon}'\hat{\epsilon} &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) = Y'Y - \hat{\beta}'X'Y + \hat{\beta}'(X'X\hat{\beta} - X'Y) \\
&= Y'Y - \hat{\beta}'X'Y = Y'Y - \hat{\beta}'(X'X\hat{\beta} = Y'(I - P)Y \\
&= \text{ sum of squares of residuals (RSS) } = \sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2
\end{aligned}$$

Example. Find least squares estimate of θ_1 and θ_2 in the following:
$y_1 = \theta_1 + \theta_2 + \epsilon_1$
$y_2 = \theta_1 - \theta_2 + \epsilon_2$
$y_3 = \theta_1 + 2\theta_2 + \epsilon_3$

Obtain X and β by writing it in the vector-matrix formulation:

$$\begin{aligned}
\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} &= \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix}, \text{ i.e.,} \\
Y &= X\beta + \epsilon.
\end{aligned}$$

Then, noting that

$$\begin{aligned}
X'X &= \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix}, \\
(X'X)^{-1} &= \frac{1}{14} \begin{pmatrix} 6 & -2 \\ -2 & 3 \end{pmatrix}
\end{aligned}$$

we obtain

$$
\begin{aligned}
\hat{\beta} &= \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = (X'X)^{-1}X'Y \\
&= \frac{1}{14}\begin{pmatrix} 6 & -2 \\ -2 & 3 \end{pmatrix}\begin{pmatrix} y_1 + y_2 + y_3 \\ y_1 - y_2 + 2y_3 \end{pmatrix} \\
&= \frac{1}{14}\begin{pmatrix} 6(y_1 + y_2 + y_3) - 2(y_1 - y_2 + 2y_3) \\ -2(y_1 + y_2 + y_3) + 3(y_1 - y_2 + 2y_3) \end{pmatrix} \\
&= \frac{1}{14}\begin{pmatrix} 4y_1 + 8y_2 + 2y_3 \\ y_1 - 5y_2 + 4y_3 \end{pmatrix} = \begin{pmatrix} \frac{2}{7}y_1 + \frac{4}{7}y_2 - \frac{1}{7}y_3 \\ \frac{1}{14}y_1 - \frac{5}{14}y_2 + \frac{2}{7}y_3 \end{pmatrix}, \\
\hat{\epsilon}'\hat{\epsilon} &= Y'Y - \hat{\beta}'X'Y = (y_1^2 + y_2^2 + y_3^2) - \frac{1}{14}(4y_1 + 8y_2 + 2y_3)(y_1 + y_2 + y_3) \\
&\quad -\frac{1}{14}(y_1 - 5y_2 + 4y_3)(y_1 - y_2 + 2y_3).
\end{aligned}
$$

Theorem. $P = X(X'X)^{-1}X'$ is symmetric idempotent, being the projection matrix onto $\mathcal{M}_C(X)$. $\text{Rank}(P) = \text{Rank}(X) = p$. $I - P$ is the orthogonal projection matrix. $\text{Rank}(I - P) = n - p$ and $(I - P)X = 0$.

The case of $\text{Rank}(X) = r < p$ will be discussed later.

An alternative derivation of $\hat{\beta}$:

$$
\begin{aligned}
(Y - X\beta)'(Y - X\beta) &= (Y - X\hat{\beta} + X\hat{\beta} - X\beta)'(Y - X\hat{\beta} + X\hat{\beta} - X\beta) \\
&= (Y - X\hat{\beta})'(Y - X\hat{\beta}) + (X\hat{\beta} - X\beta)'(X\hat{\beta} - X\beta) \\
&\quad +2(X\hat{\beta} - X\beta)'(Y - X\hat{\beta}) \\
&= (Y - X\hat{\beta})'(Y - X\hat{\beta}) + (\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta),
\end{aligned}
$$

since

$$(X\hat{\beta} - X\beta)'(Y - X\hat{\beta}) = (\hat{\beta} - \beta)'(X'Y - X'X\hat{\beta}) = 0.$$

Therefore,

$$(Y - X\beta)'(Y - X\beta) \geq (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

with equality iff $\hat{\beta} - \beta = 0$ since $X'X$ is p.d.