Steps you have an overdetermined system

1. your basic equation is
\[ \text{error}_i + A X_i^4 + B X_i^7 + C X_i + D Y_i = R_i \quad i = 1 \text{ to } m \]

2. This can be written in matrix form as
\[ [R] = [X] \beta + [E] \]

where
\[ [R] = [F] \quad [\beta] = \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} \quad [X] = \begin{bmatrix} x_1^4 & x_1^7 & x_1 & y_1 \\ x_2^4 & x_2^7 & x_2 & y_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_m^4 & x_m^7 & x_m & y_m \end{bmatrix} \quad [E] = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix} \]

3. Solve for errors
\[ [E] = [R] - [X] \beta \]

4. Square the errors
\[ = ([R]^T - \beta^T [X]^T) ([R] - [X] \beta) \]
\[ = [R]^T [R] - 2 \beta^T [X]^T [R] + \beta^T [X]^T [X] \beta \]

4. Minimize the squared errors by taking
\[ \frac{\partial ([E]^T [E])}{\partial \beta} = 0 \]
and replacing \([\beta]\) with \([b]\) because now we call \([b]\) the best estimates for \([\beta]\) in a least square sense.
4. Continued 

\[ \frac{\partial (\epsilon^T \epsilon)}{\partial \beta} = \frac{\partial (R^T R)}{\partial \beta} + \frac{\partial (2 \beta^T x^T R)}{\partial \beta} + \frac{\partial (\beta^T x^T x \beta)}{\partial \beta} \]

\[ = 0 - 2 [x^T R] + 2 [x^T x] \beta + [x^T x] \beta \]

\[ = -2 [x^T R] + 2 [x^T x] \beta = 0 \]

Solve for \[ \beta \]

\[ \Rightarrow \beta = (x^T x)^{-1} x^T R \]

5. Conclusion

Your best estimate for

\[ \hat{\beta} = \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} \]

in a least squares sense is

\[ \hat{\beta} = \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = (x^T x)^{-1} x^T R \]

Reference:


See chapter 2