Lehrstuhl für Systemverfahrenstechnik
Prof. Alexander Mitsos, PhD

Institut für Geometrie und Praktische Mathematik
Prof. Dr. A. Reusken

Model-based Estimation Methods, SS 2016

Additional Exercises (June 15, 2016)

Parameter Estimation

Solution of Bonus Problem 6: Maximum Likelihood vs. Weighted Least-Squares

$$\begin{aligned}
\hat{\Theta}_{ML} &= \arg\max_{\bar{\Theta}} L(\bar{\Theta}) \\
&= \arg\max_{\bar{\Theta}} \prod_{i=1}^{m} p(\tilde{y}_i|\bar{\Theta}, u_i) \\
&= \arg\max_{\bar{\Theta}} \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(\tilde{y}_i-\bar{y}_i)^2}{2\sigma_i^2}} \\
&= \arg\max_{\bar{\Theta}} (\prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma_i})(\prod_{i=1}^{m} e^{-\frac{(\tilde{y}_i-\bar{y}_i)^2}{2\sigma_i^2}})
\end{aligned}$$

The multiplication with a positive constant does not affect $\hat{\Theta}_{ML}$, therefore we can drop $\prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma_i}$.

$$\Rightarrow \hat{\Theta}_{ML} = \arg\max_{\bar{\Theta}} (\prod_{i=1}^{m} e^{-\frac{(\tilde{y}_i-\bar{y}_i)^2}{2\sigma_i^2}})$$

Since $e^a e^b = e^{a+b}$, we can rewrite $\hat{\Theta}_{ML}$ as

$$\hat{\Theta}_{ML} = \arg\max_{\bar{\Theta}} e^{-\sum_{i=1}^{m} \frac{(\tilde{y}_i-\bar{y}_i)^2}{2\sigma_i^2}}$$

For monotonically increasing functions g, it holds $\max_x(f(x)) = \max_x(g(f(x)))$. Since the logarithm function is monotonically increasing, it holds

$$\hat{\Theta}_{ML} = \arg\max_{\bar{\Theta}} \log(e^{-\sum_{i=1}^{m} \frac{(\tilde{y}_i-\bar{y}_i)^2}{2\sigma_i^2}})$$

Since $\log(e^{f(x)}) = f(x)$, it follows

$$\hat{\Theta}_{ML} = \arg\max_{\bar{\Theta}} (-\sum_{i=1}^{m} \frac{(\tilde{y}_i-\bar{y}_i)^2}{2\sigma_i^2})$$

Since $\max_x(-f(x)) = \min_x(f(x))$, we get

$$\begin{aligned}
\hat{\Theta}_{ML} &= \arg\min_{\bar{\Theta}}\left(\sum_{i=1}^{m} \frac{(\tilde{y}_i - \bar{y}_i)^2}{2\sigma_i^2}\right) \\
&= \arg\min_{\bar{\Theta}}\left(\sum_{i=1}^{m} \frac{(\tilde{y}_i - \bar{y}_i)^2}{\sigma_i^2}\right)
\end{aligned}$$

If we define $W = \mathrm{diag}(\sigma_1^{-1}, \ldots, \sigma_m^{-1})$, we can rewrite the sum of squares as vector-matrix-multiplication:

$$\begin{aligned}
\hat{\Theta}_{ML} &= \arg\min_{\bar{\Theta}}((\tilde{y} - \bar{y})^T W^T W(\tilde{y} - \bar{y})) \\
&= \arg\min_{\bar{\Theta}}((W\tilde{y} - W\bar{y})^T (W\tilde{y} - W\bar{y}))
\end{aligned}$$

Similar to the least-square method we define a matrix X with $X_{i,j} = x_j(u_i)$. Thus, we can rewrite $\bar{y} = X\bar{\Theta}$. For the maximum likelihood estimate we get

$$\begin{aligned}
\hat{\Theta}_{ML} &= \arg\min_{\bar{\Theta}}((W\tilde{y} - WX\bar{\Theta})^T (W\tilde{y} - WX\bar{\Theta})) \\
&= \arg\min_{\bar{\Theta}}((\tilde{y}_W - X_W\bar{\Theta})^T (\tilde{y}_W - X_W\bar{\Theta})) \\
&= \arg\min_{\bar{\Theta}} \|\tilde{y}_W - X_W\bar{\Theta}\|_2^2 \\
&= \hat{\Theta}_{WLS}
\end{aligned}$$

For models, which are linear in parameters, the maximum likelihood estimate $\hat{\Theta}_{ML}$ and the weighted least-square estimate $\hat{\Theta}_{WLS}$ are identical, if the weight matrix $W = \mathrm{diag}(\sigma_1^{-1}, \ldots, \sigma_m^{-1})$ is used in the WLS method.