Differentiating $f(\theta)$ with respect to θ_j and setting the result equal to 0 produce

$$\sum_{i=1}^{p} x_{ij}y_i = \sum_{i=1}^{p}\sum_{k=1}^{q} x_{ij}x_{ik}\theta_k.$$

If we let y denote the column vector with entries y_i and X denote the matrix with entry x_{ij} in row i and column j, these q normal equations can be written in vector form as

$$X^t y = X^t X\theta$$

and solved as

$$\hat{\theta} = (X^t X)^{-1}X^t y.$$

In the method of least absolute deviation regression, we replace $f(\theta)$ by

$$h(\theta) = \sum_{i=1}^{p}\Big|y_i - \sum_{j=1}^{q} x_{ij}\theta_j\Big|.$$

Traditionally, one simplifies this expression by defining the residual

$$r_i(\theta) = y_i - \sum_{j=1}^{q} x_{ij}\theta_j.$$

We are now faced with minimizing a nondifferentiable function. Fortunately, the MM algorithm can be implemented by exploiting the convexity of the function $-\sqrt{u}$ in inequality (3.2). Because

$$-\sqrt{u} \geq -\sqrt{u^n} - \frac{u-u^n}{2\sqrt{u^n}},$$

we find that

$$\begin{aligned} h(\theta) &= \sum_{i=1}^{p}\sqrt{r_i(\theta)^2} \\ &\leq h(\theta^n) + \frac{1}{2}\sum_{i=1}^{p}\frac{r_i^2(\theta) - r_i^2(\theta^n)}{\sqrt{r_i^2(\theta^n)}} \\ &= g(\theta \mid \theta^n). \end{aligned}$$

Minimizing $g(\theta \mid \theta^n)$ is accomplished by minimizing the weighted sum of squares

$$\sum_{i=1}^{p} w_i(\theta^n) r_i(\theta)^2$$