ECEn 671: Mathematics of Signals and Systems

Randal W. Beard

Brigham Young University

September 1, 2023

Section 1

Levenberg-Marquardt Optimization

Nonlinear Least Squares

The downside of GN is that the matrix $J^\top(x)J(x)$ may be ill-conditions at some states x.
For the general nonlinear least squares problem, we have

$$\frac{\partial \frac{1}{2}\mathbf{r}^\top(x)\mathbf{r}(x)}{\partial x} = \frac{\partial \mathbf{r}^\top}{\partial x}(x)\mathbf{r}(x) = \mathbf{J}^\top(x)\mathbf{r}(x).$$

Therefore we have

$$\begin{aligned}\text{Gradient Descent} \quad & x^{[k+1]} = x^{[k]} - \alpha \mathbf{J}^\top(x^{[k]})\mathbf{r}(x^{[k]}) \\ \text{Gauss-Newton} \quad & x^{[k+1]} = x^{[k]} - \left(\mathbf{J}^\top(x^{[k]})\mathbf{J}(x^{[k]})\right)^{-1}\mathbf{J}^\top(x^{[k]})\mathbf{r}(x^{[k]}).\end{aligned}$$

Note that there is no inverse for Gradient Descent, but it may converge slowly, even for linear residuals.

Nonlinear Least Squares

The Levenberg-Marquardt (LM) iteration is a combination of gradient descent and Gauss-Newton:

$$x^{[k+1]} = x^{[k]} - \left(\lambda I + \mathbf{J}^\top(x^{[k]})\mathbf{J}(x^{[k]})\right)^{-1} \mathbf{J}^\top(x^{[k]})\mathbf{r}(x^{[k]}),$$

where $\lambda = 1/\alpha$.
Note that $\lambda I + \mathbf{J}^\top \mathbf{J}$ is guaranteed to be full rank and well conditioned for large λ.
Standard practice:

- For the first iteration make λ large (e.g., $\approx 10^4$)
- If squared error decreases, decrease λ for next iteration (e.g., by half).
- If squared error increases, increase λ for next iteration (e.g., by 2x).

Weighted Nonlinear Least Squares

If $W = W^\top > 0$ is a weighting matrix, then

$$\min_{x \in \mathbb{R}^n} \quad \frac{1}{2}\mathbf{r}^\top(x) W \mathbf{r}(x).$$

results in

$$\begin{aligned}
&\text{(GD)} \quad x^{[k+1]} = x^{[k]} - \lambda^{-1} \mathbf{J}^\top W \mathbf{r}\big|_{x^{[k]}} \\
&\text{(GN)} \quad x^{[k+1]} = x^{[k]} - \left(\mathbf{J}^\top W \mathbf{J}\right)^{-1} \mathbf{J}^\top W \mathbf{r}\big|_{x^{[k]}} \\
&\text{(LM)} \quad x^{[k+1]} = x^{[k]} - \left(\lambda I + \mathbf{J}^\top W \mathbf{J}\right)^{-1} \mathbf{J}^\top W \mathbf{r}\big|_{x^{[k]}}.
\end{aligned}$$