ECEn 671: Mathematics of Signals and Systems

Randal W. Beard

Brigham Young University

September 1, 2023

Section 1

Gauss-Newton Optimization

Least Squares as a Gradient Descent Problem

Consider the least squares problem

$$\min_{x \in \mathbb{R}^n} \quad \|Ax - b\|_2^2$$

where $A \in \mathbb{R}^{m \times n}$ is tall. We know that the solution is

$$x^* = (A^\top A)^{-1} A^\top b.$$

Can we pose this as a gradient descent problem?

Least Squares as a Gradient Descent Problem

Define the residual as

$$\mathbf{r}(x) = \begin{pmatrix} r_1(x) \\ \vdots \\ r_m(x) \end{pmatrix} = Ax - b$$

and define the sum-of-squares error as

$$\begin{aligned} S(x) &= \frac{1}{2}\mathbf{r}^\top(x)\mathbf{r}(x) \\ &= \frac{1}{2}\sum_{j=1}^{m} r_j^2(x) \\ &= \frac{1}{2}(Ax-b)^\top(Ax-b) \\ &= \frac{1}{2}\|Ax-b\|_2^2 . \end{aligned}$$

The least squares problem is to find x that minimizes $S(x)$.

Least Squares as a Gradient Descent Problem

The gradient of S is given by

$$\frac{\partial S}{\partial x} = \frac{\partial \mathbf{r}}{\partial x}^{\top}(x)\mathbf{r}(x)$$
$$= A^{\top}(Ax - b) = A^{\top}Ax - A^{\top}b.$$

So the gradient descent algorithm gives

$$x^{[k+1]} = x^{[k]} - \alpha\left(A^{\top}Ax^{[k]} - A^{\top}b\right)$$

In general, we might allow $\alpha > 0$ to be a positive definite matrix $\mathscr{A} > 0$:

$$x^{[k+1]} = x^{[k]} - \mathscr{A}\left(A^{\top}Ax^{[k]} - A^{\top}b\right).$$

Least Squares as a Gradient Descent Problem

Selecting

$$\mathscr{A} = (A^\top A)^{-1}$$

gives

$$\begin{aligned} x^{[k+1]} &= x^{[k]} - (A^\top A)^{-1}\left(A^\top A x^{[k]} - A^\top b\right) \\ &= x^{[k]} - (A^\top A)^{-1}(A^\top A)x^{[k]} + (A^\top A)^{-1}A^\top b \\ &= (A^\top A)^{-1}A^\top b, \end{aligned}$$

which is the optimal solution.
Noting that $A = \frac{\partial \mathbf{r}}{\partial x}$, we have shown that the iteration

$$x^{[k+1]} = x^{[k]} - \left(\frac{\partial \mathbf{r}}{\partial x}^\top(x^{[k]})\frac{\partial \mathbf{r}}{\partial x}(x^{[k]})\right)^{-1}\frac{\partial \mathbf{r}}{\partial x}^\top(x^{[k]})\mathbf{r}(x^{[k]})$$

converges to the optimal in <u>one</u> step when $\mathbf{r}(x) = Ax - b$.

Nonlinear Least Squares

Let $r_j(x)$, $j = 1, \ldots, m$ be a general set of residual function to be minimized. In other words, suppose we wish to solve

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

Let $\mathbf{J}(x) \triangleq \frac{\partial \mathbf{r}}{\partial x}(x)$. Then the <u>Gauss-Newton</u> (GN) iteration algorithm is given by

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

We know that the GN method converges in one step for the linear least squares problem.