6.2 Nonlinear least squares

First consider a simple least squares problem:

$$\min_{\boldsymbol{x}} F(\boldsymbol{x}) = \frac{1}{2}\|f(\boldsymbol{x})\|_2^2. \tag{6.24}$$

Among them, the status variable is $\boldsymbol{x} \in \mathbb{R}^n$, and f is any scalar nonlinear function $f(\boldsymbol{x}) : \mathbb{R}^n \mapsto \mathbb{R}$. Note that the coefficient $\frac{1}{2}$ here is not important, some literature have this coefficient and some not. It will not affect the subsequent conclusions. Obviously, if f is a mathematically simple function, then the problem can be solved in analytical form. Let the derivative of the objective function be zero, and then find the optimal value of $\boldsymbol{x}$, just like finding the extreme value of a scalar function:

$$\frac{\mathrm{d}F}{\mathrm{d}\boldsymbol{x}} = \mathbf{0}. \tag{6.25}$$

We reach the minimum, maximum, or saddle points by solving this equation (or, intuitively, by letting the derivative be zero). But is this equation easy to solve? Well, it depends on the form of the derivative function of f. If f is just a simple linear function, then the problem is only a simple linear least squares problem. Still, some derivative functions may be complicated in form, making the equation difficult to solve. Solving this equation requires us to know the **global property** of the objective function, which is usually not possible. For the least squares problem that is inconvenient to solve directly, we can use **iterated methods** to start from an initial value and continuously update the current estimations to reduce the objective function. The specific steps can be listed as follows:

1. Give an initial value $\boldsymbol{x}_0$.
2. For k-th iteration, we find an incremental value of $\Delta\boldsymbol{x}_k$, such that the object function $\|f(\boldsymbol{x}_k + \Delta\boldsymbol{x}_k)\|_2^2$ reaches a smaller value.
3. If $\Delta\boldsymbol{x}_k$ is small enough, stop the algorithm.
4. Otherwise, let $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \Delta\boldsymbol{x}_k$ and return to step 2.

Now things get much simpler. We turn the problem of solving **the derivative function equals zero** into a problem of **looking for decreasing increments** $\Delta\boldsymbol{x}_k$. We will see that since the objective function can be linearly approximated at the current estimation, the increment calculation will be simpler *. When the function decreases until the increment becomes very small, it is considered that the algorithm converges, and the objective function reaches a minimum value. In this process, the problem is how to find the increment at each iteration point, which is a local problem. We only need to be concerned about the local properties of f at the iteration value rather than the global properties. Such methods are widely used in optimization, machine learning, and other fields.

Next, we examine how to find this increment $\Delta\boldsymbol{x}_k$. This part of knowledge belongs to the field of numerical optimization. Let's take q quick look at some of the widely used results.

* Linear cases are always the easiest ones.

1. First and Second Gradient Method

Consider the k-th iteration: we want to find Δx_k,

$\Rightarrow$ Taylor expansion (intuitive way):

$$F(x_k + \Delta x_k) \simeq F(x_k) + \underbrace{J(x_k)}_{\text{Jacobian}}{}^T \Delta x_k + \frac{1}{2} \Delta x_k^T \underbrace{H(x_k)}_{\text{Hessian}} \Delta x_k$$

In the simplest way, if we only keep the first-order one, then taking the increment at the minus gradient direction will ensure that the function decrease:

$$\Delta x^* = -J(x_k)$$

Or, we can choose to keep the second step information:

$$\Delta x^* = \arg\min \underbrace{\left(F(x) + J(x)^T \Delta x + \frac{1}{2} \Delta x^T H \Delta x \right)}$$

only contains the zero-order, first-order and quadratic terms of Δx.

finding the derivative of Δx and setting it to zero:

$$\Rightarrow \quad J + H\Delta x = 0 \Rightarrow H\Delta x = -J$$

(Newton's method)

(We omit the subscript k to simplify the notation

$\Rightarrow$ Δx means Δx_k)

2. The Gauss - Newton Method (simplest)

Idea: Carrying out a first-order Taylor expansion of $f(x)$ (note that it's $f(x)$ rather than $F(x)$):

$$f(x+\Delta x) \simeq f(x) + J(x)^T \Delta x$$

In order to find Δx such that $\|f(x+\Delta x)\|^2$ reach the minimum we need to solve a linear least square problem:

$$\Delta x^* = \arg\min_{\Delta x} \frac{1}{2}\|f(x) + J(x)^T \Delta x\|^2$$

expanding the square term of the objective function:

$$\frac{1}{2}\|f(x) + J(x)^T\Delta x\|^2 = \frac{1}{2}(f(x) + J(x)^T\Delta x)^T(f(x) + J(x)^T\Delta x)$$

$$= \frac{1}{2}(\|f(x)\|_2^2 + 2f(x)J(x)^T\Delta x + \Delta x^T J(x) J(x)^T \Delta x)$$

Find the derivative of the above formula of Δx:

$$J(x)f(x) + J(x)J^T(x)\Delta x = 0$$

We obtain:

$$\underbrace{J(x)J^T(x)}_{H(x)}\Delta x = \underbrace{-J(x)f(x)}_{g(x)}$$

$$\Rightarrow \quad H\Delta x = g$$

---

1. Set the initial value as x_0
2. For k-th iteration, calculate the Jacobian $J(x_k)$ and the residual $f(x_k)$
3. Solve the normal equation: $H\Delta x_k = g$
4. If Δx_k small enough, stop

   Otherwise, let $x_{k+1} = x_k + \Delta x_k$