1. Minimization

Optimization, either minimization or maximization of functions, is a close relation of root-finding, but it has some fundamental differences. There is a sense in which they are the same problem, because finding a stationary point of $f(x)$ is like finding a root of $\mathrm{d}f/\mathrm{d}x$. Nevertheless, they do differ (particularly in multiple dimensions, due fundamentally to the fact that the gradient of a function has special properties helpful to finding its minimum).

In the game of optimization, the most important categorization is between *convex* and *non-convex* problems. Convex problems are one with a single optimum — e.g. a single minimum. So if you find a minimum, you have found the minimum. Nonconvex problems can have multiple optima — i.e. more than one local minimum, only one of which is a global minimum (unless there's a tie!).

We start with the simplest example of a convex problem, that of a quadratic optimization. We've already seen this problem in the context of SVD, but let's connect it explicitly to the idea of minimization.

2. Minimization of quadratic functions

If you have a one-dimensional parabola, it has exactly one extreme. We will write the parabola as:

$$f(x) = \alpha x^2 + \beta x + \gamma \tag{1}$$

In this case, what is the extremum of the function?

You can easily find this extreme analytically of course: leads to

$$\begin{aligned} f'(x) = 2\alpha x + \beta &= 0 \\ x &= -\frac{\beta}{2\alpha} \end{aligned} \tag{2}$$

Typically, we will be handed a function, not given an explicit set of quadratic parameters (and we'll see later we actually want to use the quadratic form as an approximation to more general functions).

How many function evaluations do I need to do to determine the parameters of the quadratic?

Three, because there are three unknowns. If you choose points a, b, and c, these evaluations give you the equations:

$$\alpha a^2 + \beta a + \gamma = f(a)$$

$$\begin{aligned} \alpha b^2 + \beta b + \gamma &= f(b) \\ \alpha c^2 + \beta c + \gamma &= f(c) \end{aligned} \tag{3}$$

or:

$$\begin{pmatrix} a^2 & a & 1 \\ b^2 & b & 1 \\ c^2 & c & 1 \end{pmatrix} \cdot \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} f(a) \\ f(b) \\ f(c) \end{pmatrix} = \tag{4}$$

If a, b, and c are distinct, this matrix will be nonsingular.

In fact, one can work through this matrix to find α, β, nd γ (you don't really use γ, but you do need to calculate it) and thus the minimum (this is the same as Brent's method). Analytically you get the result:

$$x = b - \frac{1}{2} \frac{(b-a)^2 \left[f(b) - f(c)\right] - (b-c)^2 \left[f(b) - f(a)\right]}{(b-a) \left[f(b) - f(c)\right] - (b-c) \left[f(b) - f(a)\right]} \tag{5}$$

This method works great if the function is actually quadratic (not something else, and not linear). But is very rare for a function to be known to be quadratic, and not have the ability to calculate its derivative directly. The real use of this method is as a way to iterate twoards a solution.

3. Golden Section Search

For functions about which you know nothing, there is a more fool-proof method, akin to bisection, which is *golden section* search.

Imagine you start with a bracketing set of points a, b, and c, such that $f(b) < f(a)$ and $f(b) < f(c)$. Define the fractional position of b as:

$$w = \frac{b-a}{c-a} \tag{6}$$

Now we will pick a next trial point x. This will define some new triplet, depending on whether it is higher or lower than $f(b)$. Let's call the fraction position beyond b:

$$z = \frac{x-b}{c-a} \tag{7}$$

If the new triplet is the first three points, it has fractional length $w+z$. If it is the last three points, it has fractional length $1-w$. It seems reasonable to want these to be the same:

$$w + z = 1 - w \quad \rightarrow \quad z = 1 - 2w \tag{8}$$

which leaves each length as $1-w$.

If you are applying this iteratively, the fractional position of x within the new bracket should be the same as b was in the old bracket (since the latter was created by the same process):

$$\frac{z}{1-w} = w \tag{9}$$

and then plugging in for z, and solving for w we can find:

$$w = \frac{3-\sqrt{5}}{2} \approx 0.382 \tag{10}$$

This is the *golden section*.

This method will continually bracket the minimum further and further down.

4. Brent's one-dimensional minimization

A workable one-dimensional minimization routine combines these two methods, and is called Brent's Method. Like the root-finding Brent, it uses parabolic approximations, but it keeps track of a bracketing interval, and under certain conditions it reverts to golden section search.

These conditions are that:

- The parabolic step falls outside the bracketing interval.
- The parabolic step is greater than the step before last (!).

You shall, for your homework, implement this method.