Non-Linear Optimization without Constraints

One-dimensional rules for optimization

Taylor Series

Searching for ‘zeros’

Multi-dimensional rules

Optimization Overview

Variables: \( x = (x_1, x_2, ..., x_N) \)

Objective: \( \min f(x) \)

Subject to Constraints: \( c_i(x) = 0, i \in E \)

Today, no constraints and \( f(x) \) can be non-linear.

Single Variable Unconstrained Problems

A recap of the basic calculus method of finding the minimum and/or maximum of a function is that we solve for the \( x \) that solves the equation:

\[
\frac{\partial f(x)}{\partial x} = 0
\]

and which has a positive/negative second derivative.

Example:

Say we have a 100m length of rope. What are the dimensions of the largest area rectangular field we can enclose with it?

Let \( x = \) one dimension and \( 50 - x = \) the other dimension (this amounts to substituting the length-of-rope constraint into the problem). The objective function is the area of the rectangle implied by setting \( x \). Thus the problem is stated:

\[
\max x(50 - x) = -x^2 + 50x
\]

The solution is just:

\[
\frac{\partial f(x)}{\partial x} = -2x + 50 = 0 \quad \Rightarrow \quad x = 25
\]

i.e. the solution is a square which uses all the available rope. But we want to check for sure that this is a max. Necessary and sufficient conditions are:

\( \frac{\partial f(x)}{\partial x} = 0, \quad x \in [0, 50], \quad \text{and} \quad \frac{\partial^2 f(x)}{\partial x^2} < 0 \)

Taylor Series

The Taylor expansion formula to second order is:

\[
f(x + h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x)
\]

At nearby points to \( x \), the function behaves like this.

We can see that, if we are at the max, \( f'(x) \) must be zero and \( \frac{\partial^2 f(x)}{\partial x^2} < 0 \) otherwise a small move \( h \) would get us a higher value for the function. This
demonstrates the previously stated necessary and sufficient condition for optimization.

What if we can not ”see” $f(x)$? We could try an iterative approach where we try points $x_1, x_2, \ldots$ and try to get higher and higher with each successive try.

Unless we know something about the function and can constrain the behavior of we can’t say what happens between our observed points.

However, if we are dealing with concave/convex functions we can follow this method and find the global min/max. Two important results are:

1) If $f(x)$ is convex then any local minimum is a global minimum.
2) if $f(x)$ concave then local maximums are global maximums.

→ Also need a convex set of solutions (no gaps).
Convex objective functions then are an important class of problems. If you can show a function is concave/convex then you need only find a min/max and you know you are at the global min/max.

Two search-methods to find the max/min in 1-dimension

Bisection

**Steps:**
1) Initially search points $x_1, x_2, ...$
2) Keep most interior point with $f'(x) < 0$ and most interior point with $f'(x) > 0$
3) Pick a point half way inbetween them and:
   - if $f'(x_k+1) < 0$ replace $x_{\text{max}}$
   - if $f'(x_k+1) > 0$ replace $x_{\text{min}}$
4) Repeat until desired resolution is obtained.

**Advantages:** Known # of steps until we reach the end.

**Disadvantages:** Doesn’t use all available information. Doesn’t take into account slope and curvature.

**Newton’s Method**

This method uses information on the curvature of the function but we need to be able to calculate the curvature in order for it to be feasible.

Remember Taylor’s rule (rewritten):

$$f(x_i) = f(x_i) + (x_{i+1} + x_i)f'(x) + \frac{(x_{i+1} + x_i)^2}{2} f''(x)$$

where $x_i$ is the first try and $x_{i+1}$ is the next try.

If we maximize this approximation we use both the first and second derivative information (curvature) to make a guess as to the next point to evaluate:
\[ x_{i+1} = x_i - \frac{f'(x)}{f''(x)} \]

We must have some way of calculating the second derivative or a close approximation.

This extends to multiple dimensions using the Taylor series in higher dimensions:

\[ f(x + p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x)p \]

We have very similar optimality conditions.

\[ \nabla f(x)^T = 0 \]

\[ \nabla^2 f(x) \rightarrow \text{positive semi-definite} \]

More on this later on...

3 Line Search Methods

Steepest Decent

Newton’s Method

Quasi-newton Methods

All these methods choose a direction \( p \) and move some multiple of \( p \) in that direction towards optimum.

Steepest Decent

Assuming 1-unit steps would be taken, one idea would be to make the largest possible decrease in first order.

\[ \min \frac{g_k^T p}{\|p\|} \quad \text{or} \quad \max \frac{|g_k^T p|}{\|p\|} \]

where \( g_k = \nabla f(x) \)

\( \|p\| \) is the normalization, euclidean norm. Gives us a unit vector \( \frac{p}{\|p\|} \).

○Need to select step length also

Advantages: Easy. Will converge for sure.

Disadvantages: Could be slow (hard to know which step length).

How does line search work?

Consider:

\[ \min f(x_1, x_2) = 5x_1^2 + 7x_2^2 - 3x_1x_2 \]

Let \( x_k = (2, 3) \) if \( \alpha_k = 1 \) then: \( f \)

\( p_k = (−5, −7) \)

so \( f(x_k) = 65 \) and \( f(x_k + p) = f(\frac{1}{2}, −\frac{1}{2}) = \frac{9}{4} \)

Adding \( p_k \) gives a decrease in function value as desired. In general, we need to select \( \alpha_k \) after we know \( p_k \) and \( x_k \).
The function \( f(x_k + \alpha_k p) \) is now a function of \( \alpha_k \) and we can optimize with respect to \( \alpha_k \) using any 1-dimensional method described above. In practice, it is often more efficient to just guess \( \alpha_k \).

**Newtons Method** - Similar to 1-dimensional case.

Minimization.

We can write Taylor’s series as:

\[
 f(x + p) = f(x) + p^T g(x) + \frac{1}{2} p^T G p
\]

Where \( g = \nabla f(x) \) and \( G = \nabla^2 f(x) \)

Think of this as finding the vector that optimizes the search direction by minimizing the following w.r.t. \( p \):

\[
 Q(p) = p^T g(x) + \frac{1}{2} p^T G p
\]

A stationary point of this is the solution to \( G_k p_k = g_k \) so we use \( p_k = -G_k^{-1} g_k \)

This incorporates both first and second derivative information.

A few notes:

\( \circ G_k^{-1} \) may not be easy to find, quasi-Newton methods use approximation of \( G_k^{-1} \) at each step.

\( \circ \) If derivatives are difficult to calculate then we can use "finite difference" methods.

\( \circ \) Line search methods need a termination rule. Pick the desired resolution of the solution and stop when closer than that.

\[
 \nabla f(x) \leq \epsilon
\]

and

\[
 \nabla^2 f(x) + \epsilon I \text{ is positive semi-definite.}
\]