Christopher Lum
lum@uw.edu

Lecture 10g
Numerical Optimization Algorithms: Armijo Rule

Outline

-Choosing a Step Size

-Armijo Rule

Introduction

As we saw previously, in unconstrained optimization problems, it may be difficult or impossible to solve analytically for stationary points. Therefore, we result to numerical techniques to try and find these stationary points.

The major disadvantage to the gradient descent with line minimization is that it requires parametrization of the function. What if you do not have a functional representation of the cost function (it is a look-up table of various data points)? Or what if it was discontinuous or otherwise difficult to parameterize. In this scenario, other methods for choosing the step size α^k should be investigated.

One of the most popular and effective methods for choosing the step size is called the Armijo rule.

Armijo Rule

In this rule, we have three constants

$$s > 0 \qquad \textbf{(Eq.A.1)}$$

$$\beta \in (0, 1) \qquad \textbf{(Eq.A.2)}$$

$$\sigma \in (0, 1) \qquad \textbf{(Eq.A.3)}$$

The step size using this method is

$$\alpha^k = \beta^{m_k} s \qquad \textbf{(Eq.A.4)}$$

We can immediately see why we require $s > 0$ (or else $\alpha^k \leq 0$)

Here, m_k is the first non-negative integer m for which

$$f(x^k) \geq f(x^k + \beta^m s\, d^k) - \sigma \beta^m s \nabla f(x^k)^T d^k \qquad \textbf{(Eq.A.5)}$$

Armijo Rule Algorithm

So we see for this rule, the procedure/algorithm/pseudo-code for the Armijo rule is shown below

```
function [alpha_k] = ArmijoRule(d_k,s,beta,sigma)

    1. We are given:
        -The specified descent direction $d^k$
        -Ability to evaluate cost function at any location (we need ability to compute $f(x^k)$, $f(x^{k+1})$)

    2. Choose $s > 0$ and $\beta$ and $\sigma$ in the range of (0,1).

    3. Start with $m = 0$

    4. Start while(1) loop
    5. Inside loop, check

      If $(f(x^k) \geq f(x^k + \beta^m s d^k) - \sigma \beta^m s \nabla f(x^k)^T d^k)$
        $m_k = m$;
         break out of while(1) loop
      else
         m = m + 1;
      end

    5. Compute $\alpha^k = \beta^{m_k} s$
```

So at each step of the algorithm, we can call our 'ArmijoRule' function to compute the appropriate step size.

We investigate an example of this in the homework.

Analysis

Let us analyze the Armijo rule.

We make some observations:

If d^k is a descent direction, then $\nabla f(x^k)^T d^k < 0$

$s > 0$

$\sigma \in (0, 1)$ (positive)

$\beta^m > 0$ (because $\beta \in (0, 1)$ and $m \geq 0$)

We now examine Eq.1.5. Using the immediately preceding observations, we note that in this equation, the second term (counting the minus sign) is positive. In other words

$$-\sigma\,\beta^m\,s\,\nabla f\left(x^k\right)^T d^k > 0.$$

Let us denote this quantity as the improvement factor, IF^m.

$$\text{IF}^m = -\sigma\,\beta^m\,s\,\nabla f\left(x^k\right)^T d^k > 0 \qquad\qquad \textbf{\textit{(Eq.2)}}$$

We note that the improvement factor is positive and decreases as *m* increases.

```
In[∘]:= σ = 0.5;
        β = 0.8;
        s = 50;
        gradFdotd = -2.3;
        Table[{m, -σ β^m s gradFdotd}, {m, 0, 25}] // TableForm
        Clear[σ, β, s, gradFdotd]
```

```
Out[∘]//TableForm=
0     57.5
1     46.
2     36.8
3     29.44
4     23.552
5     18.8416
6     15.0733
7     12.0586
8     9.6469
9     7.71752
10    6.17402
11    4.93921
12    3.95137
13    3.1611
14    2.52888
15    2.0231
16    1.61848
17    1.29478
18    1.03583
19    0.828662
20    0.66293
21    0.530344
22    0.424275
23    0.33942
24    0.271536
25    0.217229
```

So Eq.A.5 can be written as

$$f\left(x^k\right) \geq f\left(x^k + \beta^m\,s\,d^k\right) + \text{IF}^m \qquad\qquad \text{recall: } \alpha^k = \beta^{m_k}\,s$$

$$f\left(x^k\right) \geq f\left(x^k + \alpha^k\,d^k\right) + \text{IF}^m \qquad\qquad \text{recall: } x^{k+1} = x^k + \alpha^k\,d^k$$

$$f(x^k) \geq f(x^{k+1}) + \text{IF}^m$$

So we see that by choosing the step size via Armijo Rule ensures that the cost function value at the subsequent point $x^{k+1} = x^k + \alpha^k d^k$ will be less than the cost function value at the current point x^k by at least the positive improvement factor of $\text{IF}^m = -\sigma \beta^m s \nabla f(x^k)^T d^k$ (again, recall that $\text{IF}^m > 0 \; \forall \; m$).

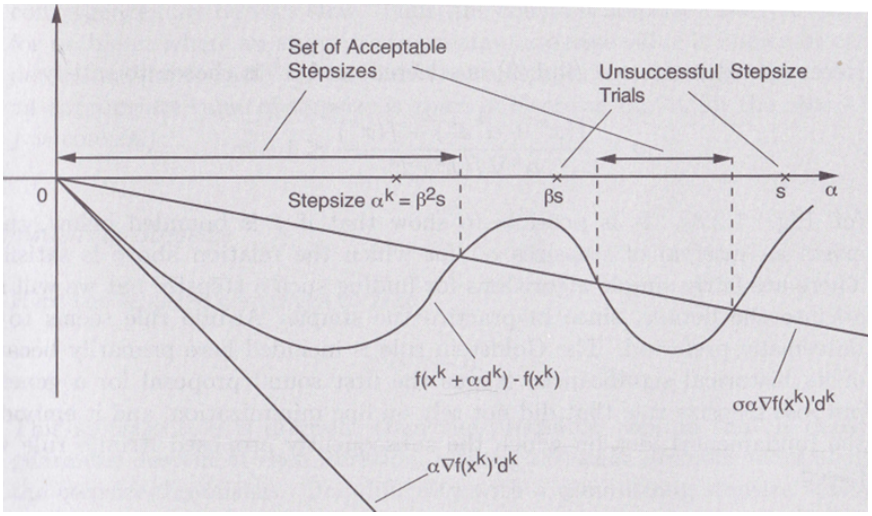
We can rewrite Eq.A.5 as (letting $\alpha = \beta^m s$)

$$f(x^k) - f(x^k + \alpha d^k) \geq -\sigma \alpha \nabla f(x^k)^T d^k$$

$$f(x^k + \alpha d^k) - f(x^k) \leq \left[\sigma \nabla f(x^k)^T d^k\right] \alpha \qquad \textbf{\textit{(Eq.A.6)}}$$

We recognize the term $\nabla f(x^k)^T d^k$ as the directional derivative of f at x^k in the direction d^k. Since d^k is a descent direction, we know that as we move away from x^k initially the function values must decrease. Therefore, $\nabla f(x^k)^T d^k < 0$. The term $\nabla f(x^k)^T d^k \alpha$ vs. α (with α on the x-axis) is therefore a line passing through the origin with a negative slope. This slope is made less negative by the coefficient σ.

Therefore, the valid values of α are those where the function $f(x^k + \alpha d^k) - f(x^k)$ is below the line defined by $\left[\sigma \nabla f(x^k)^T d^k\right] \alpha$. This explains the picture in Figure 1.2.7 of the Bertsekas text.



So now with a direction and step size, we can implement $x^{k+1} = x^k + \alpha^k d^k$ until we find a stationary point.

Now that we have a better understanding of how the algorithm executes, we can better understand what the different parameters mean and how they influence the algorithm

$s > 0$

this is the initial/largest step size to consider

$\beta \in (0, 1)$

how aggressively α changes with each Armijo Rule iteration

closer to 1 means α does not change much

closer to 0 means α has larger changes at each iteration

$\sigma \in (0, 1)$

how much improvement do we demand at each iteration of the Armijo Rule

closer to 1 means the slope of the acceptable line vs. α (with α on the x-axis) is closer to $\nabla f\left(x^k\right)^T d^k$

closer to 0 means the slope of the acceptable line vs. α (with α on the x-axis) is flatter and closer to 0 (horizontal)

Armijo Rule Execution