Gradient Descent Method

Sai Krishna Kalyan

26th November 2015

1 Introduction

The scope of this assignment is to implement descent methods for solving unconstrained optimisation problems. The initial points must lie in the domain of f and in addition subset level must be closed.

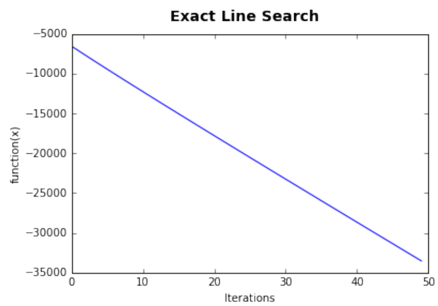
2 Descent Methods

Generally descent methods are as follows. It alternates between two steps, determining direction d, and step size t.
Given a starting point in domain f
repeat
Determine direction d.
Line Search, choose step size t, exact or backtracking line search.
Update $x = x + t * d$
until stopping criterion is satisfied.
For this assignment we will consider the following function below and minimise it using different descent methods.

$$f(x) = c^T x - \sum_{i=1}^{m} \log(b_i - a_i^T x)$$

2.1 Exact Line Search

Exact line search is a method often used in optimisation practice.Basically for this exact line search, an array of t was chosen with the following values: [.0001, .0005, .001, .005, .01, .05, .1, .5, 1, 1.5] and for each iteration, the value of x is updated by the formula $x = x + t * d$. t is obtained by minimising $f(x + s * d)$ where t is minimised with respect to s, with s being greater than or equal to 0.



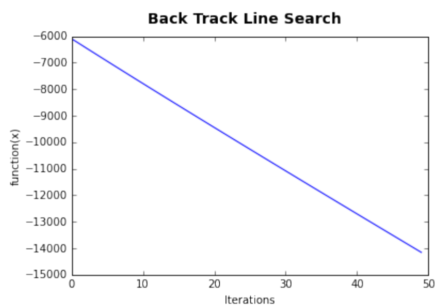
2.2 Backtrack Line Search

In this method t is reduced by factor β which is updated when condition(given below) is satisfied by $t\beta$.

$$\textbf{while } f(x + t\Delta x) > f(x) + \alpha t \nabla f(x)^T \Delta x, \quad t := \beta t.$$

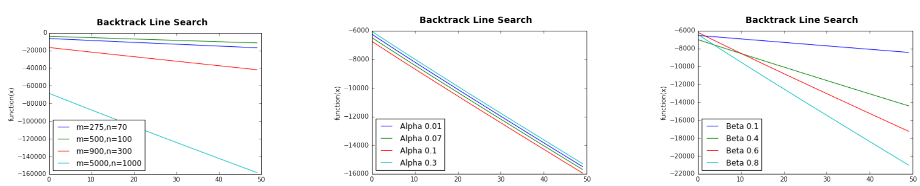
.

The plot below gives us a linear slope and took more computation time compare to exact line search



.

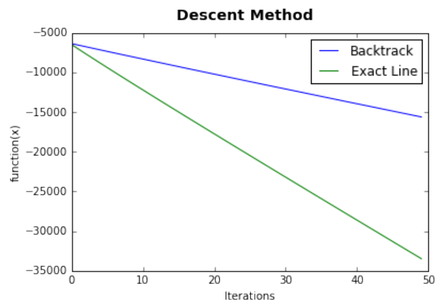
This method depends on two parameters α, β where α varies between 0.01 and 0.3 with 0.1 as default meaning that we accept a decrease in f between 1 and 30 precent of the prediction based on the linear extrapolation. β varies between 0.1 to 0.8 with 0.5 as default.



Numerical Experiments [L-R] varying Dimensions, α, β

The reason lines *function(x)* have a different starting point is because of different seeds for x.

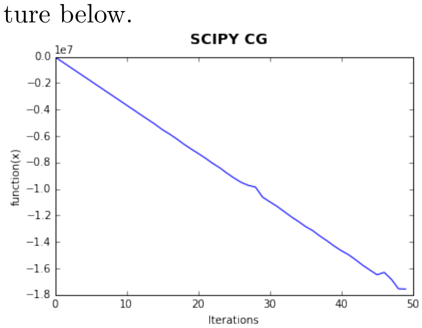
My observations mentioned below
Increasing dimensions causes longer computation cycles with steeper slopes.
Varying α does not seem have a significant impact.
Varying β gives us a steeper descent which suggest that a value of α between 0.4 to 0.6 should be good.



Plot above seems to suggest that sometimes Exact Line search might give us better convergence compared to Backtrack Line search.

3 Conjugate Gradient Algorithm

Conjugate gradient algorithm is based on work of Polak and Ribiere. Using the fmincg function in the scipy.optimize package we minimise our objective function. Ploting the objective function vs number of iterations gives us picture below.



4 Conclusions

We can make the conclusions summarised below.
The choice of backtracking parameters α,β have a noticeable but not dramatic effect on the convergence.
Exact Line search sometimes has better rate of convergence.
Generally Backtrack and Exact Line search exhibit the properties of linear convergence.