Ch.11 Quasi-Newton methods

QSnake

An Introduction to Optimization

4/21,2023

1. Ideas of Newton methods
2. Introduction
3. Approximating the Inverse Hessian
4. The Rnak One Correction Formula
5. The DFP Algorithm
6. The BFGS Algorithm

Ideas of Newton methods

Ideas of Newton methods

$$x_{k+1} = x_k - F(x_k)^{-1} g_k$$

Key idea is $1^{st} + 2^{nd}$ derivatives more efficient

Shortcoming of Newton methods

1. May have numerical problem
2. for a general non-linear function, may be no converge.

Introduction

First step: add the alpha

$$x^{(k+1)} = x_k - \alpha_k F(x^{(k)})^{-1}$$

where α_k is chosen to ensure that

$$f(x_{k+1}) < f(x_k)$$

Second step: change the second derivative

$$x_{k+1} = x_k - \alpha H_k g_k$$

where H_k is an $n \times n$ real matrix

Proposition

$$x^{k+1} = x^{(k)} - \alpha_k H_k g_k$$

where $\alpha_k = \arg\min_{\alpha \geq 0} f(x^{(k)} - \alpha H_k g_k)$

then $\alpha_k > 0$ and $f(x_{k+1}) < f(x_k)$

Approximating the Inverse Hessian

Assumption

1. $F(x)$ is constant and independent of x
2. $F(x) = Q$ for all x ($Q = Q^T$)

The Performance of Q

$$g^{(k+1)-g^{(k)}} = Q(x^{k+1} - x^{(k)}) \equiv \Delta g^{(k)} = Q\Delta x^{(k)}$$

$$Q^{-1}\Delta g^{(i)} = \Delta x^{(i)} \quad 0 \leq i \leq k$$

The Performance of H

Therefore, we also impose the requirement that the approximation H_{k+1} of the Hessian satisfy:

$$H_{k+1}\Delta g^{(i)} = \Delta x^{(i)} \quad 0 \leq i \leq k$$

This set of equation can be represented as

$$Q^{-1}[\Delta g^{(0)}, \cdots, \Delta g^{(n-1)}] = [\Delta x^{(0)}, \cdots, \Delta x^{(n-1)}]$$

Quasi-Newton methods

$$x^{k+1} = x^k + \alpha_k d^k$$

where $d^{(k)} = -H_k g^{(k)}$
$H_0, \cdots$ are symmetric and satisfy $H_{k+1}\Delta g^{(i)} = \Delta x^{(i)}, \quad 0 \le i \le k$
$\alpha_k = \arg\min_{\alpha \ge 0} f(x^k + \alpha d^{(k)})$

Also Q-conjugate

$If \alpha_i \neq 0,\ 0 \leq i \leq k$, then $d^{(0)}, \cdots, d^{k+1}$ are Q-conjugate.

The Rnak One Correction Formula

Key Idea

$$H_{k+1} = H_k + a_k z^{(k)} z^{(k)T}$$

where $a_k \in \mathbb{R} \& z^{(k)} \in \mathbb{R}^n$

Final Form

$$H_{k+1} = H_k + \frac{(\Delta x^{(k)} - H_k \Delta g^{(k)})(\Delta x^{(k)} - H_k \Delta g^{(k)})^T}{\Delta g^{(k)T}(\Delta x^{(k)} - H_k \Delta g^{(k)})}$$

Rank one Algorithm

1. Set $k = 0$, select $x^{(0)}$ and a real symmetric positive definite H_0
2. If $g^{(k)} = 0$, stop; else, $d^{(k)} = -H_k g^{(k)}$
3. Compute $x^{(k+1)}$ by $x^{(k)} + \alpha_k d^{(k)}$
4. Compute H_{k+1}
5. continue to step 2.

Problem

1. H_{k+1} may not be positive definite
2. If $\Delta g^{(k)}(\Delta x^{(k)} - H_k \Delta g^{(k)})$ is close to zero, then there may be numerical problems in evaluating H_{k+1}

The DFP Algorithm

DFD Algorithm

1. Set $k = 0$; select $x^{(0)}$ and a real symmetric positive definite H_0
2. $d^{(k)} = -H_k g^{(k)}$
3. Compute $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)} (\alpha_k = \arg\min_{\alpha \geq 0} f(x^{(k)} - \alpha d^{(k)}))$
4. Compute H_{k+1} where

$$H_{k+1} = H_k + \frac{\Delta x^{(k)} \Delta x^{(k)T}}{\Delta x^{(k)T} \Delta g^{(k)}} - \frac{[H_k \Delta g^{(k)}][H_k \Delta g^{(k)}]^T}{\Delta g^{(k)T} H_k \Delta g^{(k)}}$$

5. continue to step 2

Property

In the DFP algorithm applied to the quadratic with Hessian $Q = Q^T$, we have $H_{k+1}\Delta g^{(i)} = \Delta x^{(i)}$

Problem

In the case of larger non-quadratic problems,the algorithm has the tendency of sometimes getting 'stuck'.

The BFGS Algorithm

Ideas

Approx Q not Q^{-1}

Ideas

$$\Delta g^{(i)} = B_{k+1} \Delta x^{(i)}, \quad 0 \leq i \leq k$$

where B_k be our estimate of Q at kth step

$$H_{k+1}^{DFP} = H_k + \frac{\Delta x^{(k)} \Delta x^{(k)T}}{\Delta x^{(k)T} \Delta g^{(k)}} - \frac{[H_k \Delta g^{(k)}][H_k \Delta g^{(k)}]^T}{\Delta g^{(k)T} H_k \Delta g^{(k)}}$$

$$B_{k+1} = B_k + \frac{\Delta g^{(k)} \Delta g^{(k)T}}{\Delta g^{(k)T} \Delta x^{(k)}} - \frac{[B_k \Delta x^{(k)}][B_k \Delta x^{(k)}]^T}{\Delta x^{(k)T} B_k \Delta x^{(k)}}$$

so $H_{k+1}^{BFGS} = (B_{k+1})^{-1}$(use Shermon-Morrison formular)