Week 3, Lecture 1

01 February 2024 14:25

In this lecture we continue to look at the BFGS formula for the descent direction. We investigate the circumstances in which the BFGS method produces a positive-definite approximation of the Hessian at each iteration.

This is based on **Chapter 4** of the typed notes.

BFGS Algorithm (Ch. 4)

Hessian Matrix

$$[B(\underline{x}_k)]_{ij} = \left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{\underline{x}_k}$$

<u>Aim</u>: Avoid computing B_k^{-1} at each iteration ($O(n^3)$)

Approximate:

$$\underbrace{\underline{y}_k}_{\underline{\nabla} f(\underline{x}_{k+1}) - \underline{\nabla} f(\underline{x}_k)} \simeq \underbrace{B(\underline{x}_k)}_{= B_{k+1}} \underbrace{\underline{s}_k}_{\underline{x}_{k+1} - \underline{x}_k}$$

Hence:

$$\underline{y}_k = B_{k+1} \underline{s}_k$$

- n equations
- $n \times n$ unknowns

BFGS solves this eqn in an approximate sense.

Two vector quantities:

$$\underline{y}_k \,, \qquad B_k \underline{s}_k$$

Our approximate sl^{n} to Eqn is built out of $\underline{y}_k$ and $B_k \underline{s}_k$:

$$\begin{cases} B_{k+1} = B_k + \alpha\, \underline{y}_k \underline{y}_k^T \\ \qquad\qquad + \beta\, \underline{\underline{(B_k \underline{s}_k)}}(B_k \underline{s}_k)^T . \end{cases}$$

Here, $\underline{y}_k \underline{y}_k^T$ is the outer product:

$$(n \times 1)(1 \times n) = n \times n .$$

$$[\underline{y}_k \underline{y}_k^T]_{ij} = (y_k)_i (y_k)_j .$$

We require that the approximation should satisfy the secant equation:

$$\underline{y}_k = B_{k+1} \underline{s}_k$$

We have found (lecture notes) the **general formula**:

General formula:

$$B_{k+1} = B_k + \frac{\underline{y}_k \underline{y}_k^T}{\langle \underline{y}_k, \underline{s}_k \rangle} - \frac{B_k \underline{s}_k \underline{s}_k^T B_k^T}{\langle \underline{s}_k, B_k^T \underline{s}_k \rangle} \qquad (1)$$

In this lecture, we look in depth at the properties of the matrices B_k and B_{k+1} in Equation (1).

Positive-Definite Property (§ 4.7.1)

Theorem: Provided B_0 is positive definite and symmetric, and provided the curvature condition:

$$\langle y_k, s_k \rangle > 0$$

is satisfied at each iteration, then the BFGS method produces a symmetric positive-definite approximation to the Hessian at each iteration.

EXAM PROOF

The proof is by induction on *k*. We start with the symmetry property:

Symmetry:

$$B_{k+1} = B_k + \underbrace{\alpha\, y_k y_k^T + \beta\, (B_k s_k)(B_k s_k)^T}$$

$$= B_k + \text{Symmetric Matrices}$$

By assumption, B_0 is symmetric, so by mathematical induction, B_k is symmetric for all $k \in \{0, 1, 2, \dots\}$.

We next look at the positive-definite property:

To show: $\langle \xi, B_{k+1} \xi \rangle > 0 \;\forall\, \xi \neq 0$.

FIRST TWO TERMS : Δ

$$\langle \underline{\xi}, B_{k+1}\underline{\xi}\rangle = \langle \underline{\xi}, B_k\underline{\xi}\rangle - \frac{\langle \underline{\xi}, B_k \underline{s}_k \underline{s}_k^T B_k^T \underline{\xi}\rangle}{\langle \underline{s}_k, B_k^T \underline{s}_k\rangle} + \frac{\langle \underline{\xi}, \underline{y}_k \underline{y}_k^T \underline{\xi}\rangle}{\langle \underline{y}_k, \underline{s}_k\rangle}$$

Look at the **last term**, highlighted in yellow:

Last term:

$$\frac{\langle \underline{\xi}, \underline{y}_k \underline{y}_k^T \underline{\xi}\rangle}{\langle \underline{y}_k, \underline{s}_k\rangle} = \frac{\langle \underline{\xi}, \underline{y}_k\rangle^2}{\underbrace{\langle \underline{y}_k, \underline{s}_k\rangle}_{>0,\ \text{curvature condition}}} > 0.$$

So it suffices to show that the first two terms are positive definite.

$$\Delta = \langle \underline{\xi}, B_k \underline{\xi}\rangle - \frac{\langle \underline{\xi}, B_k \underline{s}_k \underline{s}_k^T B_k \underline{\xi}\rangle}{\langle \underline{s}_k, B_k \underline{s}_k\rangle}.$$

Since B_k is symmetric positive-definite,

$$B_k \underline{u}_i = \lambda_i \underline{u}_i, \qquad \lambda_i > 0.$$

$$\underline{\xi} = \sum_i \xi_i \underline{u}_i, \quad \xi_i = \langle \underline{\xi}, \underline{u}_i \rangle$$

$$\underline{s}_k = \sum_i \sigma_i \underline{u}_i, \quad \sigma_i = \langle \underline{s}_k, \underline{u}_i \rangle.$$

$$\langle \underline{\xi}, B_k \underline{\xi} \rangle = \sum_i \lambda_i \xi_i^2$$

Sub these relations back into the expression for Δ

$$\Delta = \langle \underline{\xi}, B_k \underline{\xi} \rangle - \frac{\langle \underline{\xi}, B_k \underline{s}_k \underline{s}_k^T B_k \underline{\xi} \rangle}{\langle \underline{s}_k, B_k \underline{s}_k \rangle}$$

$$= \sum_i \lambda_i \xi_i^2 - \frac{\langle \sum_i \xi_i \underline{u}_i, B_k \underline{s}_k \underline{s}_k^T B_k \sum_j \xi_j \underline{u}_j \rangle}{\sum_i \lambda_i \sigma_i^2}$$

Re-arrange the sums:

$$\Delta = \sum_i \lambda_i \xi_i^2 - \frac{\sum_i \sum_j \xi_i \xi_j \langle \underline{u}_i B_k \underline{s}_k \underline{s}_k^T B_k \underline{u}_j \rangle}{\sum \lambda_i \sigma_i^2}$$

Move B_k and use the fact that it's symmetric:

$$\Delta = \sum_i \lambda_i \xi_i^2 - \frac{\sum_i \sum_j \xi_i \xi_j \langle \overset{= B_k}{B_k^T} \underline{u}_i, \underline{S}_k \underline{S}_h^T \underline{B_k u_j} \rangle}{\sum \lambda_i \sigma_i^2}$$

Use $B_k \underline{u}_i = \lambda_i \underline{u}_i$

$$\Delta = \sum_i \lambda_i \xi_i^2 - \frac{\sum_i \sum_j \xi_i \xi_j \lambda_i \lambda_j \langle \underline{u}_i, \underline{S}_k \underline{S}_h^T \underline{u}_j \rangle}{\sum_i \lambda_i \sigma_i^2}$$

Use the following **key fact**:

$$\langle \underline{u}_i, \underline{S}_k \underline{S}_h^T \underline{u}_j \rangle = \sigma_i \cdot \sigma_j$$

Hence:

$$\Delta = \sum_i \lambda_i \xi_i^2 - \frac{\sum_i \sum_j \xi_i \xi_j \lambda_i \lambda_j \cdot \sigma_i \cdot \sigma_j}{\sum \lambda_i \sigma_i^2}$$

Introduce $\underline{X}$ and $\underline{Y}$ such that:

$$X_i = \sqrt{\lambda_i}\, \xi_i \,, \quad Y_i = \sqrt{\lambda_i}\, \sigma_i \,.$$

Hence:

$$\Delta = \frac{\left(\sum_i \lambda_i \xi_i^2\right)\left(\sum_i \lambda_i \sigma_i^2\right) - \sum_i \xi_i \lambda_i \sigma_i \sum_j \xi_j \lambda_j \sigma_j}{\sum \lambda_i \sigma_i^2}$$

Common denominator

$$= \frac{\left(\sum_i X_i^2\right)\left(\sum_i Y_i^2\right) - \left(\sum_i X_i Y_i\right)\left(\sum_j X_j Y_j\right)}{\sum_i Y_i^2}$$

Re-write this as:

$$\Delta = \frac{\|\underline{X}\|_2^2 \|\underline{Y}\|_2^2 - (\underline{X} \cdot \underline{Y})^2}{\|\underline{Y}\|_2^2}$$

Hence:

$$\Delta \geq 0 \text{, by C.S.}$$

So

$$\langle \underline{\xi}, B_{k+1} \xi \rangle = \Delta + \frac{\overbrace{\langle \underline{\xi}, y_k \rangle}^{2}}{\langle y_k, \underline{s}_k \rangle} > 0.$$

So B_{k+1} is positive-definite.

Hence:

B_k is p. def. $\Rightarrow$ B_{k+1} is p. def.

Since B_0 is p. def. by mathematical induction, B_k is p. def. for all $k \in \{0, 1, 2, \dots\}$. $\blacksquare$