Newton Methods

Hoonyoung Jeong
Department of Energy Resources Engineering
Seoul National University

Review

- Explain the steepest gradient method intuitively
- Explain a step size
- Present stop criteria
- What is the disadvantage of the steepest gradient method?
- Explain the conjugate gradient method by comparing with the steepest gradient method

Key Questions

- Explain the meaning of Taylor series and how to utilize
- Explain the Newton method intuitively
- Derive the iteration equation of the Newton method using the 1^{st} and 2^{nd} order Taylor series
- Present the disadvatages of the Newton method
- What are the quasi-Newton methods?
- Compare DFP and BFGS
- Derive the equation of the Gauss-Newton method
- Present and explain remedies for gradient-based optimization

Approximation using Taylor Series

- $f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \cdots$
- Meaning and How to utilize?
  - ✓Small $|x-a|$ ➔ $|x-a|^n \approx 0$
  - ✓$f(x)$ can be approximated using 1st order, 2nd order, or 3rd order, …
  - ✓For $a - \epsilon < x < a + \epsilon$
    - 1st order: $f(x) = f(a) + f'(a)(x-a)$
    - 2nd order: $f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2$
  - ✓Approximate $f(x)$ for $x \approx a$ once you know $f(a)$, $f'(a)$, $f''(a)$

Overview for Newton Method

- Newton method.pdf

What is Newton Method?

- Newton method for $f(x) = 0$
  - ✓Approximate $f(x)$ using 1st order Taylor Series
  - ✓Find a solution for $f(x) = 0$
  - ✓Repeat
- Newton method for $f'(x) = 0$
  - ✓Approximate $f'(x)$ using 1st order Taylor Series or Approximate $f(x)$ using 2nd order Taylor Series
  - ✓Find a solution for $f'(x) = 0$
  - ✓Repeat

Newton Method for $f(x) = 0$

- Finding x satisfying $f(x) = 0$
- 1st order Taylor series approximation (TSA)
- Calculate x_{k+1} at x_k iteratively using 1st order TSA
- $f(x_{k+1}) = f(x_k) + f'(x_k)(x_{k+1} - x_k)$
- $0 = f(x_k) + f'(x_k)(x_{k+1} - x_k)$
- $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$

Newton Method for $f'(x) = 0$

- Finding x satisfying $f'(x) = 0$
- 1st order Taylor series approximation (TSA)
- Calculate x_{k+1} at x_k iteratively using 1st order TSA of $f'(x)$
- $f'(x_{k+1}) = f'(x_k) + f''(x_k)(x_{k+1} - x_k)$
- $0 = f'(x_k) + f''(x_k)(x_{k+1} - x_k)$
- $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$

Newton Method for $\nabla f(x) = 0$

- Multivariate
- $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$
- $x_{k+1} = x_k - \frac{1}{H(x_k)} \nabla f(x_k)$
- $x_{k+1} = x_k - H(x_k)^{-1} \nabla f(x_k)$
  - ✓ $H(x_k)^{-1}$ is a step size in terms of the gradient descent method

Disadvantages of Newton Method for $f'(x) = 0$

- $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$
- What if $f(x)$ is concave down?
  - ✓ Wrong direction
- What if $f''(x_k) \approx 0$?
  - ✓ Divergence
- Numerical calculation of $f''(x_k)$ might be unstable
  - ✓ $f'(x_k) = \frac{f(x_k+\Delta x) - f(x_k)}{\Delta x}$
  - ✓ $f''(x_k) = \frac{f'(x_k+\Delta x) - f'(x_k)}{\Delta x}$
- Disadvantages of Newton Method for $\nabla f(x) = 0$
  - ✓ What if $f(x)$ is concave down?
    - Wrong direction
  - ✓ What if $H(x_k)$ is singular?
    - $H(x_k)^{-1}$ doesn't exist, divergence
  - ✓ Numerical calculation of $H(x_k)$ might be unstable

We can’t use Newton Method?

- No worry! We have good numerical methods
  - ✓ Wrong direction
    - $x_{k+1}=x_k-\frac{f'(x_k)}{|f''(x_k)|}$
    - $x_{k+1}=x_k-|H(x_k)|^{-1}\nabla f(x_k)$
    - Switch eigenvalues of H to positive signs
  - ✓ Singularity of $H(x_k)$
    - Reduce singularity
    - $x_{k+1}=x_k-(|H(x_k)|+\mu_k I)^{-1}\nabla f(x_k)$
    - Increase μ_k: make the quadratic curve sharper and the step size smaller
  - ✓ Trust region
    - $|H(x_k)|^{-1}$ can be considered a step size of GDM
    - What if the step size ($|H(x_k)|^{-1}$) is inaccurate?
    - Limit the range of the next solution, $|x-x_k|=r_k$
    - $x_{k+1}=x_k-\lambda_k H(x_k)^{-1}\nabla f(x_k)$
  - ✓ $x_{k+1}=x_k-\lambda_k(|H(x_k)|+\mu_k I)^{-1}\nabla f(x_k)$

Quasi-Newton Methods

- Numerical approximation of Hessian matrix using function values
- What if $H(x_k)$ is positive-definite, non-singular, and stable?
- DFP algorithm (Davidon-Fletcher-Powell)
  - ✓Approximate the inverse matrix of a Hessian matrix
  - ✓Gaurantee positive definiteness
- BFGS algorithm (Broyden–Fletcher–Goldfarb–Shanno)
  - ✓Best numerical calculation of a Hessian matrix
  - ✓Gaurantee positive definiteness
  - ✓Self-correctness: Even if the approximation of H is inaccurate, the approximation of H is corrected in a few steps

DFP vs. BFGS

$$x_{k+1} = x_k + \lambda_k d_k$$

DFP

$$d_k = -A_k c_k$$

$$c_k = \nabla f(x_k)$$

$$A_{k+1} = A_k + B_k + C_k$$

$$A_0 = I, \quad B_k = \frac{s_k s_k^T}{s_k \cdot y_k}, \quad C_k = \frac{-z_k z_k^T}{y_k \cdot z_k}$$

$$s_k = \lambda_k d_k \qquad z_k = A_k y_k$$

$$y_k = c_{k+1} - c_k$$

BFGS

$$d_k = -H_k^{-1} c_k$$

$$c_k = \nabla f(x_k)$$

$$H_{k+1} = H_k + D_k + E_k$$

$$H_0 = I, \quad D_k = \frac{y_k y_k^T}{y_k \cdot s_k}, \quad E_k = \frac{c_k c_k^T}{c_k \cdot d_k}$$

$$s_k = \lambda_k d_k$$

$$y_k = c_{k+1} - c_k$$

Gauss-Newton, Levenberg, Levenberg-Marguardt Methods

- Gauss-Newton_Levenberg_LM.pdf

Summary of Gradient-based Optimization Methods

- Review of Gradient-based Optimization Methods.pdf