Steepest Gradient Method
Conjugate Gradient Method

Hoonyoung Jeong
Department of Energy Resources Engineering
Seoul National University

Review

- What is the formula of ∇f?
- What is the physical meaning of ∇f?
- What is the formula of a Jacobian matrix?
- How is a Jacobian matrix utilized?
- What is the formula of a Hessian matrix?
- How is a Hessian matrix utilized?

Key Questions

- 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

Comparison Intuitive Solution to Steepest Gradient Method

The example is mathematically expressed as:

- ✓ $f(\mathrm{x})$: elevation at location $\mathrm{x} = [x \quad y]$
- ✓ Find x maximizing $f(\mathrm{x})$

1) Find the steepest ascent direction by taking one step to each direction at the current location
   Steepest ascent direction
   = Tangential direction at the current location
   = $\nabla f(\mathrm{x}_k) = \left[\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y}\right]$
   = Direction locally maximizing f(x) or Locally steepest ascent direction

2) Take steps along the steepest direction

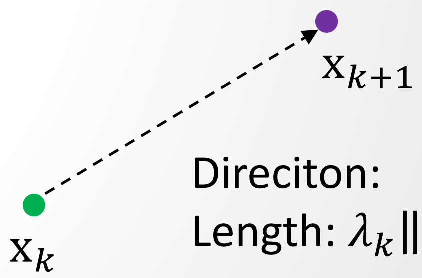
$$\mathrm{x}_{k+1} = \mathrm{x}_k + \lambda_k \nabla f(\mathrm{x}_k)$$

   k: iteration
   λ_k: how many steps to the steepest direction

3) Stop if you meet uphill

4) Repeat 1) – 3) until you reach a peak

$$\mathrm{x}_{k+1} - \mathrm{x}_k = \lambda_k \nabla f(\mathrm{x}_k)$$



Direciton:
Length: $\lambda_k \|\nabla f(\mathrm{x}_k)\|$

Example of Gradient Descent Method (1)

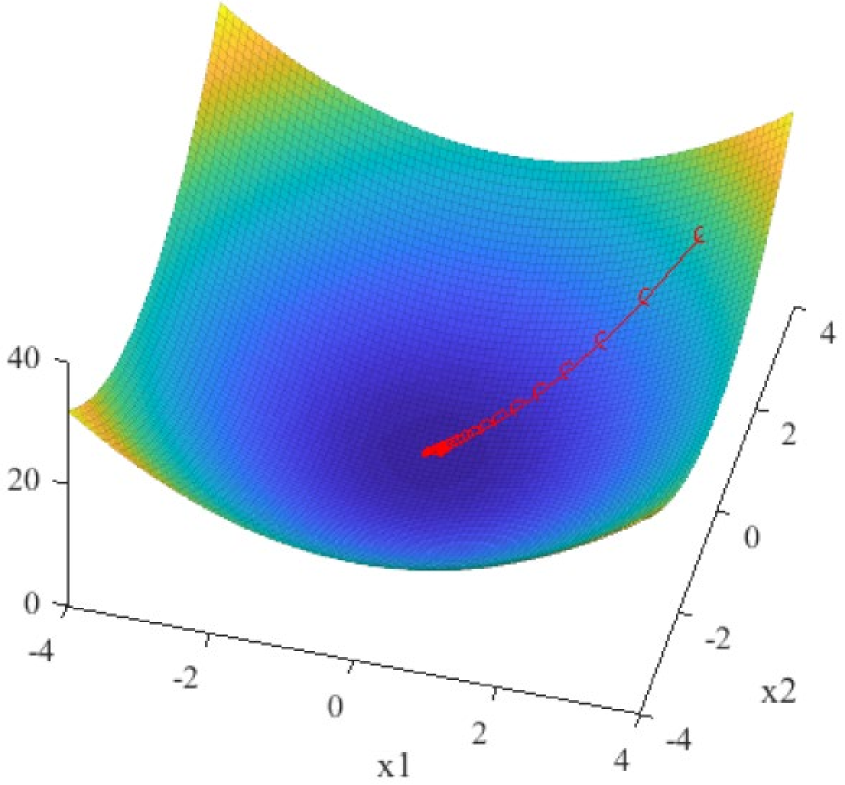
- $x_{k+1} = x_k - \lambda_k \nabla f(x_k)$
- $f = x^4$
- ∇f
- $x_0 = 2$
- λ_k =0.02 (constant)
- $x_1 = x_0 - \lambda_0 \nabla f(x_0)$=?
- $x_2 = x_1 - \lambda_1 \nabla f(x_1)$=?
- $x_3 = x_2 - \lambda_2 \nabla f(x_2)$=?
- Repeat for λ_k =0.1 (constant)

Example of Gradient Descent Method (2)

- See gradient_descent_method.mlx

Disadvantage of Gradient Descent Method

- Slow near local minima or maxima
  - ✓The 2-norm of a gradient vector becomes small near local minima or maxima

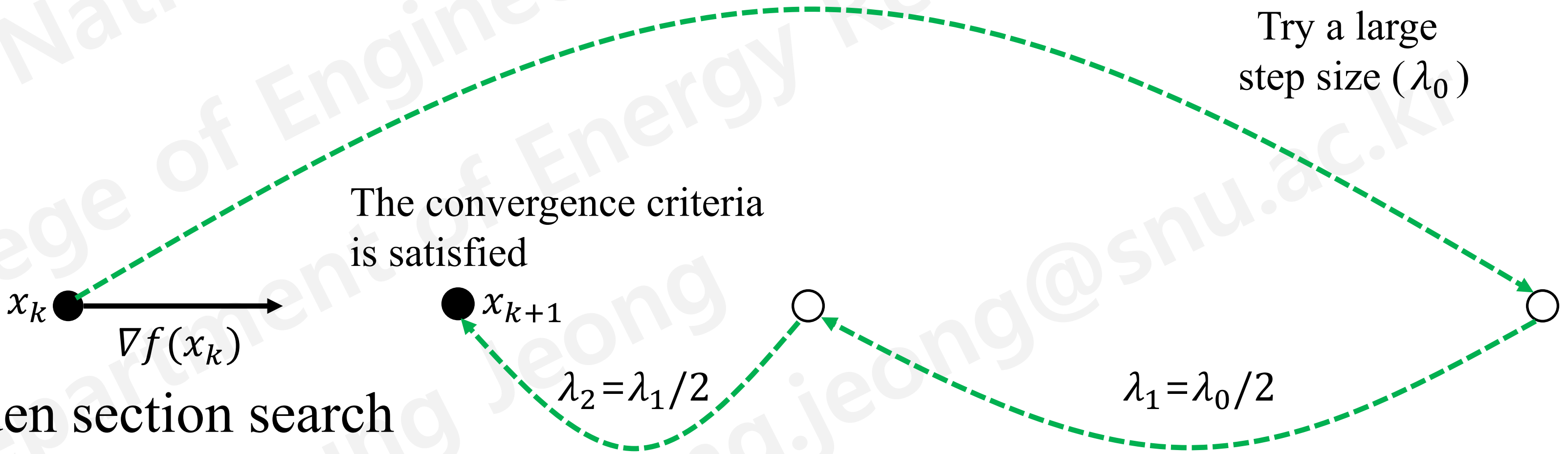


Step Size

- Step length / Learning rate called in machine learning
- What are the advantages of large and small step sizes?
  - ✓Large: fast
  - ✓Small: stable
- What are the disadvantages of large and small step sizes?
  - ✓Large: unstable
  - ✓Small: slow
- Step size
  - ✓Should be appropriate
  - ✓But, different for different problems

How Determine a Step Size?

- Line search methods
  - ✓Backtracking line search



  - ✓Golden section search
  - ✓Secant method
- No implementation, No need to know details
  - ✓Why? Matlab already has good line search methods

Stop (Convergence) Criteria

- 2-norm of ∇f
  - ✓ $|\nabla f(x_k)| < \epsilon_{|\nabla f|}$, ex) 1e-5
- Relative change of f
  - ✓ $\frac{|f(x_{k+1})-f(x_k)|}{|f(x_{k+1})|} < \epsilon_f$, ex) 1e-5
- Relative change of x
  - ✓ $max\left(\frac{|x_{k+1,i}-x_{k,i}|}{x_{k+1,i}}\right) < \epsilon_x \; for \; i = 1, \ldots, N_x$, ex) 1e-10
- Maximum number of iterations
- Maximum number of f evaluations

Conjugate Gradient Method

- A modification of the gradient descent method
- Add a scaled direction to the search direction
- The scaled direction is computed using the previous search direction, a nd the current and previous gradients
- More stable convergence because the previous search direction is cons idered (less oscillations)

SGM vs. CGM

Steepest Gradient Method

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

$c_k = \nabla f(x_k)$

$d_k = -c_k$

search direction

Ascent $d_k = -c_k$

Conjugate Gradient Method

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

$c_k = \nabla f(x_k)$

$d_k = -c_k + \left(\frac{|c_k|}{|c_{k-1}|} \right)^2 d_{k-1}$

Ascent $d_k = c_k + \left(\frac{|c_k|}{|c_{k-1}|} \right)^2 d_{k-1}$

Example of SGM and CGM

- Manual calculation
  - ✓Example_of_SGM_and_CGM.pdf
- MATLAB
  - ✓Example_of_SGM_and_CGM.mlx