08. 일차 미분 이용한 방법들

(First-Order Methods)

↓

경사 하강법 (Gradient Descent)

접합 경사하강법 (Conjugate Gradient)

ㄴ공액, 켤레

경사 하강법 (Gradient Descent)

- 경사도 : 목적 함수의 1차 미분
- 하강 방향 : d (경사가 큰 방향)
  - d로 이동할 경우, 목적 함수의 최소값으로 계획점 이동
  - 목적 함수가 2차 함수의 형태이며, 단계값 충분히 작고 경사도가 '0'이 아니라는 가정.

    → 변화 심하지 X
- 경사 하강 방향 : $\nabla f(x)$
- $g^{(k)} = \nabla f(x^{(k)})$, k번째 반복의 계획점 $x^{(k)}$
- $d^{(k)} = -\dfrac{g^{(k)}}{\|g^{(k)}\|}$: 표준화 (normalize)
  - $d^{(k)}$의 길이를 1로 만든다.
  - $x^T = [x_1, x_2] \Rightarrow \|x\| = \sqrt{x_1^2 + x_2^2}$ (norm)
- (장) gradient Descent 방법은 자료주어진 (surrogate) 목적함수 최소화 문제에 多 사용

  방향만 결정되니까 구조 간단, 계산 안정적
- (단) 반복수 多. 목적 함수 형태에 따라 최적값 얻지 못함

ex) $f(x) = x_1 x_2^2$, $x^{(k)} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$\nabla f(x)^T = [x_2^2, \ 2x_1x_2]$

$\nabla f(x) = \begin{bmatrix} -4 \\ -4 \end{bmatrix} = g^{(k)}$

$\|g^{(k)}\| = \sqrt{(-4)^2 + (-4)^2} = 4\sqrt{2}$

$d^{(k)} = \dfrac{g^{(k)T}}{\|g^{(k)}\|} = \left[-\dfrac{1}{\sqrt{2}}, \ -\dfrac{1}{\sqrt{2}}\right]$

gradient descent 의 특징 : 방향 vec는 직교!

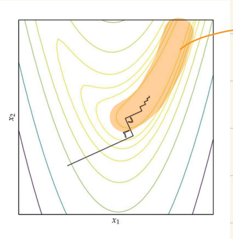
- 직교성 (orthogonality)
  - 직교 : $x^T y = 0$ ← 가지고 있는 원소 수 같아야 내적 가능
  - $\alpha^{(k)} = \arg\min_{\alpha} f(x^{(k)} + \alpha d^{(k)})$
  - $\nabla f(x^{(k)} + \alpha d^{(k)})^T d^{(k)} = 0$
  - $d^{(k+1)} = -\dfrac{\nabla f(x^{(k)} + \alpha d^{(k)})}{\|\nabla f(x^{(k)} + \alpha d^{(k)})\|}$ ← gradient
  - $\nabla f(x^{(k)} + \alpha d^{(k)})^T d^{(k)} = 0$ ← α에 의한 목적함수 최소값이 되려면 gradient = 0

    $\rightarrow d^{(k+1)T} d^{(k)} = -\dfrac{\nabla f(x^{(k)} + \alpha d^{(k)})^T}{\|\nabla f(x^{(k)} + \alpha d^{(k)})\|} d^{(k)} = 0$

  ∴ 두 방향 vec는 서로 직교성 만족!

  ↳ 두 vec의 사이값 90°

ex) Rosenbrock banana function 방향 vec



narrow valley (협곡)

: 천천히 줄어들며 최소값에 완만히 이동!

↓

경사하강법은 narrow valley 에서 계산값 불안정해 질 수 있다!

접합 경사 하강법 (conjugate Gradient)

최소값이 정해지기 위해서 목적함수의 조건

- FONC : first-order necessary condition

$$f(x^*) \le f(x+hy) \Leftrightarrow f(x+hy) - f(x^*) \ge 0$$

1차 미분해 테일러 전개 값 이용

테일러 2차 전개

- SONC : second-order necessary condition

$$f(x^*+hy) = f(x^*) + h \underbrace{\nabla f(x^*)^T y} + \frac{1}{2}h^2 y^T \nabla^2 f(x^*) y + O(h^3)$$

y는 방향 vec, gradient 값 → 이 값이 0일 때 최소점

$$\Rightarrow \frac{1}{2}h^2 y^T \nabla^2 f(x^*) y = f(x+hy) - f(x^*) \ge 0$$

⇒ FONC, SONC 합침.

- IDEA

2차 형식 : quadratic

$$\underset{x}{\text{minimize}} \; f(x) = \boxed{\frac{1}{2}x^T A x} + b^T x + c$$

- f(x)는 유일한 국소 최소점
- A : 대칭(symmetric)인 양 정치 (Positive definite) 행렬
- n차원 이차함수는 n번 반복에 의해 최적화.

- 행렬 A에 의한 $d^{(i)}$와 $d^{(j)}$는 서로 접합 (켤레, 공액)

$: d^{(i)T} A d^{(j)} = 0, \quad \forall i, j$ → 반드시 직교X ↝ 2차 형식 이니까!

방향 vec

- 접합 경사 하강 알고리즘

- 접합 방향 : 이전 하강 방향과 경사도 이용해 결정

1) $d^{(1)} = -g^{(1)}$

2) $x^{(2)} = x^{(1)} + \alpha^{(1)} d^{(1)}$: $\alpha^{(1)}$은 선탐색으로 결정

↳ line search Method
backtracking
strong backtracking

k+1) $d^{(k+1)} = -g^{(k+1)} + \beta^{(k)} d^{(k)}$

$\beta^{(k)}$도 찾아내야함 (식으로 유도!)

* 2차 함수 이면 α는 계산식 존재

- 이차 함수의 α

α 찾는 식 찾아 반복계산 X

$$\underset{\alpha}{\text{minimize}} \; f(x+\alpha d)$$

$(x-\alpha d)^2$ 형태

$$\frac{\partial f(x+\alpha d)}{\partial \alpha} = \frac{\partial}{\partial \alpha}\left[\frac{1}{2}\underbrace{(x-\alpha d)^T A (x-\alpha d)} + b^T(x-\alpha d) + c\right]$$

$$= d^T A (x-\alpha d) + d^T b$$

$$= d^T(Ax+b) - \alpha d^T A d = 0$$

$$\therefore \alpha = -\frac{d^T(Ax+b)}{d^T A d}$$ (직접해, 식 이용해 계산)

이차 함수 아닐 경우 line search 등의 방법 이용

- β 유도

$$d^{(i)T} A d^{(j)} = 0$$

$d^{(i)}$와 $d^{(j)}$는 켤레
A는 대칭 & 양 정치 행렬

$$\Rightarrow (-g^{(k+1)} + \beta^{(k)} d^{(k)})^T A d^{(k)} = 0$$

$$\Rightarrow -g^{(k+1)T} A d^{(k)} + \beta^{(k)} d^{(k)T} A d^{(k)} = 0$$

$$\Rightarrow \frac{g^{(k+1)T} A d^{(k)}}{d^{(k)T} A d^{(k)}}$$

→ 반복 계산 X, 직접해
A행렬 파악 불가능해서 단점

- A 선택

- Fletch-Reeves (1964) : $\beta^{(k)} = \dfrac{g^{(k)T} g^{(k)}}{g^{(k-1)T} g^{(k-1)}}$

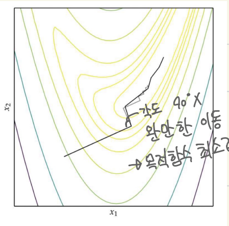
↳ 굳이 A 없어도 됨!

- Polak-Ribière (1969)

$$: \beta^{(k)} = \frac{g^{(k)T}(g^{(k)} - g^{(k-1)})}{g^{(k-1)T} g^{(k-1)}}, \quad \underbrace{\beta^{(k)}} \leftarrow \max(\beta^{(k)}, 0)$$

↳ 음수 제외
: 항상 최적값 나옴!

Rosenbrock 에서의 Conjugate Gradient



각도 90° X
완만한 이동
→ 목적함수 최소값에 빠르게 계산

↓

두 방향은 Conjugate (○)
orthogonal (X)