9장 부록

보충 설명: 9.2절

선형 탐색법의 수렴(Convergence of line search method)

선형 탐색법을 사용하는 경사 하강 알고리즘이 전역적으로 수렴하기 위한 조건은 수학적으로 Zoutendijk 정리에 의해 설명된다. 이 정리는 선형 탐색 기반 알고리즘이 수렴하기 위한 조건을 제시하며, 특히 탐색 방향 $\boldsymbol{d}_k$가 최급강하 방향(steepest descent direction)으로부터 얼마나 벗어나도 수렴할 수 있는지를 설명한다.

정리 9A.1 (Zoutendijk 정리) [^Nocedal et al. (2006)]

초기점 $\boldsymbol{x}_0 \in \mathbb{R}^n$에서 시작하여 다음과 같이 이터레이션을 정의하자:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \eta_k \boldsymbol{d}_k, \tag{1}$$

여기서:

- $\boldsymbol{d}_k$는 하강 방향 (즉, $\nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_k < 0$),
- η_k는 Wolfe 조건을 만족하는 학습률,
- $f \in C^1(\mathcal{D}, \mathbb{R})$: 1차 도함수가 연속인 함수,
- f는 아래로 유계이며, 초기값보다 함수값이 작거나 같은 점들의 집합

  $$\mathcal{L} := \{\boldsymbol{x} \in \mathcal{D} : f(\boldsymbol{x}) \leq f(\boldsymbol{x}_0)\} \tag{2}$$

  을 포함하는 열린집합 $\mathcal{D} \subset \mathbb{R}^n$ 위에서 정의됨,
- ∇f는 $\mathcal{D}$에서 립시츠 연속

이러한 조건이 모두 만족되면 다음의 Zoutendijk 조건이 성립한다:

$$\sum_{k=0}^{\infty} \cos^2 \theta_k \|\nabla f(\boldsymbol{x}_k)\|^2 < \infty, \qquad \text{(Zoutendijk condition)} \tag{3}$$

여기서:

$$\cos \theta_k := \frac{-\nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_k}{\|\nabla f(\boldsymbol{x}_k)\| \cdot \|\boldsymbol{d}_k\|}. \tag{4}$$

위 조건 (3)은 다음 극한을 유도한다:

$$\cos^2\theta_k \|\nabla f(\boldsymbol{x}_k)\|^2 \to 0. \tag{5}$$

즉, 그래디언트가 0으로 수렴하거나, 탐색 방향이 점차 최급강하 방향과 직각에 가까워진다는 것을 의미한다.

이제 추가적인 조건을 통해 좀 더 강한 결론을 도출할 수 있다. 만약 모든 k에 대해 탐색 방향이 일정 각도 이상으로 최급강하 방향과 벗어나지 않는다면, 다시 말해 양의 상수 $\delta > 0$가 존재하여

$$\cos\theta_k \geq \delta > 0, \tag{6}$$

을 만족한다면, 다음의 전역 수렴 조건이 성립한다:

$$\lim_{k\to\infty} \|\nabla f(\boldsymbol{x}_k)\| = 0. \tag{7}$$

이러한 결과는 $\boldsymbol{x}_k$가 극소점(minimizer)으로 수렴한다고 보장하는 것은 아니다. 단지 **정상점(stationary point)**으로 수렴함을 보장한다. 따라서 실제로는 정지 조건이 만족되는 점으로 수렴할 뿐, 그 점이 꼭 전역 최소점일 필요는 없다.

이제 뉴턴 또는 준뉴턴 방법을 생각하자. 이 경우, 탐색 방향은

$$\boldsymbol{d}_k = -B_k^{-1}\nabla f(\boldsymbol{x}_k) \tag{8}$$

의 형태로 주어진다. 이때 B_k는 양의 정부호 대칭행렬이며, 다음 조건을 만족한다고 하자:

$$\|B_k\| \cdot \|B_k^{-1}\| \leq M, \quad \forall k \tag{9}$$

즉, B_k의 **조건수(condition number)**가 균일하게 유계(uniformly bounded)임을 의미한다.

이 경우, 단순한 계산을 통해 다음이 성립함을 보일 수 있다:

$$\cos\theta_k \geq \frac{1}{M} > 0, \tag{10}$$

따라서 $\cos\theta_k$가 양의 하한을 가지므로 위의 수렴 조건이 적용되어 다음이 보장된다:

$$\lim_{k\to\infty} \|\nabla f(\boldsymbol{x}_k)\| = 0. \tag{11}$$

이는 뉴턴 방법 및 준뉴턴 방법이 적절한 조건 하에서 **Wolfe 조건을 만족하는 선형 탐색법**을 사용할 경우 전역 수렴성을 갖는다는 중요한 이론적 결과이다.

보충 설명: 9.3절

옵티마이저 시각화(Visualization of optimizers)

다음 그림을 통해 앞서 제시된 최적화 알고리즘이 어떻게 다르게 동작하는지 이해 할 수 있다.

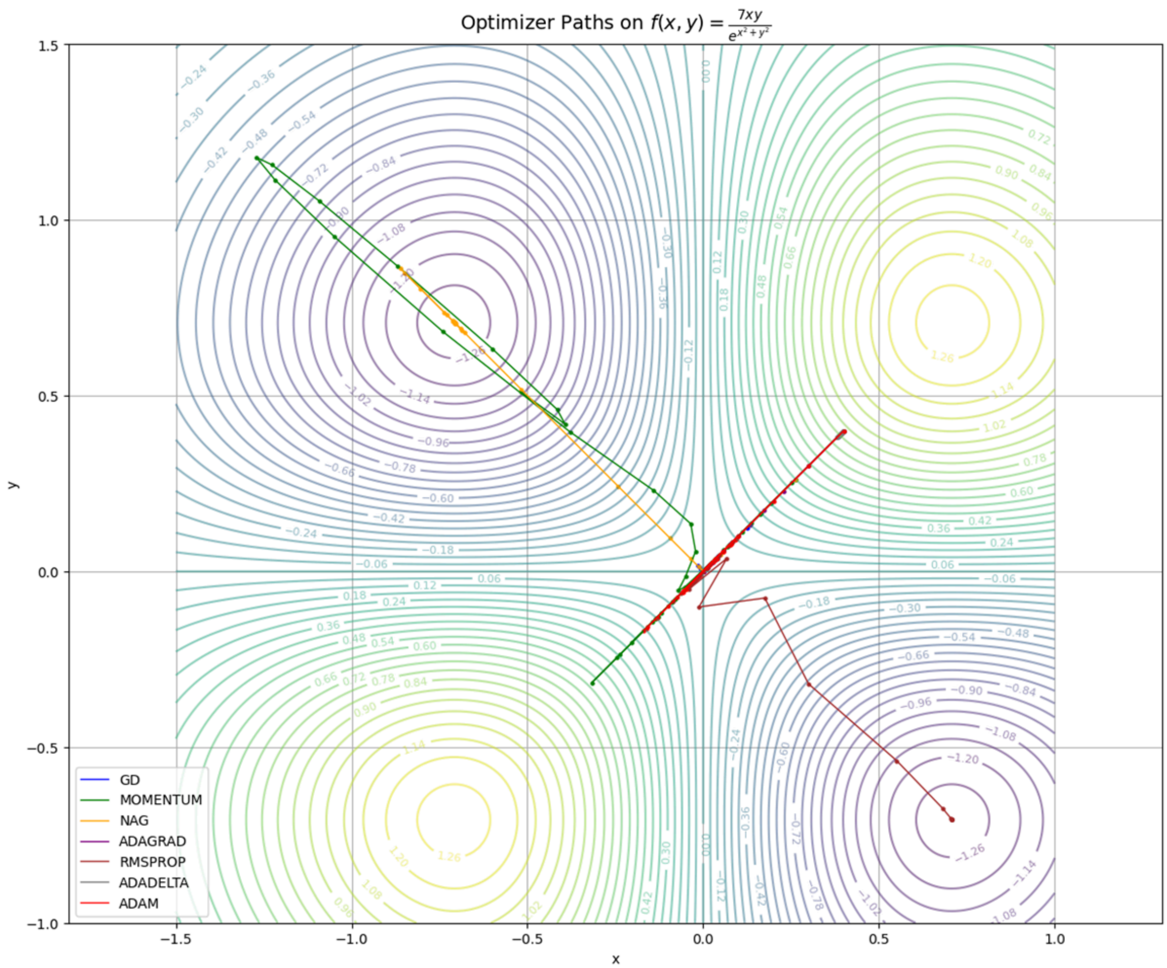

그림 9-A.1. 옵티마이저 시각화

각주